

US free markets less than free

The United States, the land of free enterprise, seems to be curiously illogical in its continuing belief that markets in some commodities should be less than free

SHOULD Eastern Europe and the Soviet Union look as confidently as they do towards the United States for a model of how the free-market system functions? Not necessarily, if some recent US economic behaviour is to be taken seriously. By now, the whole world knows how troublesome it has been for the Congress to arrive at an acceptable gasoline tax in this year's federal budget; in the end, it settled for a mere 5 cents a US gallon, making a total federal tax of 14 cents a gallon and leaving the cost of gasoline much lower than on the world's spot markets.

But there is worse. As if recognizing that gasoline in the United States will still be cheaper than it should be, the Congress has clapped a tax on motor-cars that fail to yield their drivers at least 23.5 miles for every US gallon. The result will not be a free market in gasoline consumption, but one in which patterns of consumption are skewed by contradictory and illogical forces.

That the price of gasoline should be a contentious issue is no surprise, of course. Last week's protests in Hungary, which took the form of huge traffic jams that can only have increased the cost of travelling, are a sufficient proof of that. The rest of Eastern Europe will find the going hard this winter, as the Soviet Union requires payment for its supply of oil in hard currency and as the governments of Eastern Europe begin living with the reality that they can no longer print banknotes to let their people buy the goods they need. Yet the price that angers Hungarians is roughly twice the going price of gasoline in the United States, where there is almost as vivid a sense that its price has been unjustly ramped up.

Yet there should be no doubt, in a true free market, what the price of gasoline should be. It must, for example, exceed the marginal cost, or that of producing a small extra amount of it. Otherwise there would be no reason why those who refine petroleum should continue to operate their distillation plant and catalytic crackers. This explains why gasoline prices have been rising in the past few months; uncertainty over the course of events in the Persian Gulf has raised doubts about the continuity of a large part of the world's supply of crude oil, so that those trading in this market are now less willing to part with oil they happen to own because they expect to be able to sell it for an even higher price if there should be physical shortages of crude in the next few months. This is not merely consistent with what the textbooks say, but also with the outcome of the British government's investiga-

tion of the sharp increases in the price of gasoline from British petrol pumps during September. It is true that oil companies lucky enough to control cheaper supplies can hope to profit hugely at times like these, but that is their good fortune (which may nevertheless be abated by taxation) and their chief incentive for opening up other sources of supply.

Evidently this elementary lesson is not generally appreciated in the United States. For one thing, there is general grumbling that higher gasoline prices will cause hardship to many social groups, and that prices should be kept in check for that reason. The complaint is true, but the way to take the edge off social hardship is not to shield all gasoline consumers from higher prices, but to give extra cash to those who need it. For another, in the closely related field of energy supply in general, there is a whole regulatory apparatus, at the federal and state level, designed to relate the prices charged for energy to average rather than marginal costs. This is the effect of the procedures by which the prices charged by public utilities for commodities such as natural gas are fixed with reference to the overall profitability of the supplier, for example. It is curious that the land of free enterprise should so shield people from market reality.

This is the spirit in which the 'gas-guzzler' tax that has won the favour of both parties in the Congress should be seen for what it is — as a kind of moral statement of opinion about the ownership of inefficient motor-cars rather than a contribution to the management of a quickly changing energy market or, as some claim, as a contribution to the abatement of carbon-dioxide emissions. □

Utah confusion

The time has come for decisive action by Utah's Fusion Advisory Committee.

IN the 18 months since cold fusion was first brought into the world by Stanley Pons and Martin Fleischmann, the original evidence for the phenomenon has become tenuous to the point of invisibility. The tritium evaporated and the gamma rays disappeared. Then the central claim, for excess heat generation, began to look shaky.

And now Stanley Pons himself is becoming scarce. A



flurry of newspaper stories last week hinted that Pons had "vanished" — his house sold, his telephone disconnected, his son out of school — only days before the Fusion Advisory Committee of the State of Utah was to decide whether to continue its support of the National Cold Fusion Institute (NCFI) in Salt Lake City. Among many at the University of Utah, there was hope that, in the absence of any report from Pons or Fleischmann of what had been achieved over the past year, the NCFI would be gently put to sleep.

But it was not to be. Pons, communicating to the world by fax through his chosen representative C. Gary Triggs (a North Carolina attorney well known to critics of cold fusion for his careful reading of their papers), revealed that he was in France, and keen to take a sabbatical from the University of Utah. This was news to the university, it seemed. The Fusion Advisory Committee, evidently worried about the fate of the \$5 million so far spent on the NCFI but not yet ready to write off its investment, postponed its decision until 8 November, by which time a team of external reviewers is supposed to have scrutinized — in the space of a one-day visit — NCFI's progress. There are hints that Pons will visit Salt Lake City in time for the meeting on 8 November, but in the meantime Fritz Will, the director of NCFI, has publicly chastized Pons for his lack of cooperation. This suggests one model for the slow demise of cold fusion: like other failed revolutionary movements, it will disappear in a blizzard of factional fighting, the Marxist-Leninists (so to speak) and the Leninist-Marxists each professing to the same beliefs, but unable to stomach each other's company.

There is ample opportunity to make fun of these goings-on and, in the light of the history of cold fusion, no good reason not to. But if there is a serious comment to be made of this latest episode of the soap opera, it is perhaps to lament the inability of the Utah legislature to bring the proceedings to a halt. Pons did not appear at a meeting expressly held to decide the future of the NCFI, will not communicate with the institute's director, and yet the Fusion Advisory Committee still, apparently, feels obliged to grant the itinerant chemist another period of grace. If a total rebuff is not sufficient to dispel the committee's confidence in cold fusion, what will be? □

Universities for sale?

The British government's plan to turn British universities into commercial enterprises has backfired.

No good ever seemed likely to come of the latest scheme to make British universities compete against each other for public funds (see page 3). And so it has proved. But there is a kind of justice in the way in which this half-baked scheme has blown up in the faces of the government and of its designated hit-man, the Universities Funding Council (UFC). The obvious and demeaning error is UFC's miscalculation that, after a decade of offic-

ial beastliness towards British academic institutions, the universities would do anything for money. The underlying difficulty is the British government's attempt to hide continuing central control of higher education beneath a patina of illusory autonomy.

It is a sorry tale. Three years ago, when Mr Kenneth Baker's Great Education Reform Bill was rumbling through the British Parliament, the government floated the idea that British universities would be required under the new regime to negotiate contracts with the government for the provision of what might be called educational services (as is now required of Australian universities, for example). That proposal was shouted down, but UFC promptly devised a largely equivalent scheme — that universities would make money 'bids' for the privilege of teaching specified numbers of students in predefined fields. UFC was seeking a way of teaching larger numbers of students at less than the proportionately increased cost. Helpfully, it published figures (literally called "guide prices") for the notional cost of teaching students for a year in different fields of study. The idea was that universities wishing to expand in such a field would prudently submit a lower bid. UFC's intention was then to add together all the bids (for student numbers and for costs) so as to arrive at the most economical package.

So, guess what? Universities have submitted bid prices essentially the same as the guide prices put out by UFC. Collectively, they want to teach more students, but not at lower unit cost. UFC apparently accepts defeat, although it has not yet said how it will distribute its funds among its dependants in future years. (That is where the retribution will be felt.) Apparently the informal understanding reached by university vice-chancellors that they would not depart much from the guide prices will immediately be followed by a reference of the universities to the Monopolies and Mergers Commission — the nearest thing in Britain to anti-competition legislation.

This bidding fiasco derives directly from the government's confusion about its policy on higher education. For much of the past decade, it has sought to restrict student numbers. At the beginning of the decade, fearful that the cost of student maintenance would get out of hand, it required UFC's then predecessor to threaten universities with penalties if student quotas were exceeded, with the consequence that the system ran below capacity. The decision that there should be expansion instead has come only in the past three years. That welcome conversion to good sense is calamitously late. But at no stage during this long period has the British government been willing to contemplate an arrangement in which universities would be genuinely free to decide for themselves how they would balance the needs of students and research, of diversity and specialization. What nobody appreciates is that there are many universities in Britain that would gladly trade freedom for a little money. Ironically, the government may find that its aborted bidding system, by institutionalizing the guide prices, may have made it more difficult to win economies. □

University funding plan collapses in chaos

■ Expansion plans in question

■ UFC to think again

THE UK Universities Funding Council (UFC) has plans to increase access to university education while decreasing the cost of teaching each student are in tatters after the controversial system which had required universities to bid against one another for funded student places had to be abandoned last week. Sir Edward Parkes, chairman of the Committee of Vice-Chancellors and Principals (CVCP) said that the UFC's decision leaves the universities living "from hand to mouth", unable to plan beyond the next academic year.

Under the new bidding system, universities told UFC how many students they wished to teach in each subject from 1991–92 to 1994–95, and the price at which each student could be taught. The UFC was to have assessed these bids and distributed student places for the four-year planning period (see *Nature* 342, 843; 1989).

But a letter from the UFC to all vice-chancellors, dated 23 October, says that decisions on funding after the 1991–92 academic year have been postponed, because the UFC is "disappointed by the scale of economy offered by the universities' bids". The competitive nature of the system was supposed to encourage universities to realize 'economies of scale' and bid for more students at a reduced cost per student, in line with government policy. The letter explains that universities will be informed of their teaching grants for 1991–92 in February, but decisions on funding for the rest of the planning period will have to wait until the bidding system has been reconsidered.

Most university vice-chancellors argued that decreasing teaching costs per student would threaten the quality of university education. In total, universities bid for a 19 per cent increase in students by 1994–95, with 93 per cent of these places priced at the UFC's suggested maximum 'guide prices' (based on the current average cost of teaching students in each subject area).

Even before the bids had been submitted, it was obvious that the government would not pay for university expansion at the prices being suggested by the universities (see *Nature* 345, 468; 7 June 1990). At a press conference called to respond to the UFC letter, Parkes said that the UFC seemed to have only now recognized the problem, which "the most junior lecturer" could have identified months ago. But he denied that the universities had operated a

cartel to subvert the bidding system. Although the universities rejected the UFC's views on the potential for economy, he said, they had welcomed the move away from the annual allocation of teaching grants towards a system based on four-year planning periods.

A UFC spokesman says there is "no intention to return to a year-by-year funding system", and that funding for the remainder of the four-year planning period will be addressed in "the coming

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Sir Edward Parkes — universities are left "living from hand to mouth".

He says the bidding system will be reviewed, but it is not yet clear what the UFC's approach will be. Just how the UFC will allocate funds and student places for 1991–92 is also uncertain — the government's public expenditure statement later this month is unlikely to release sufficient funds to allow even a modest expansion at the present cost per student.

Parkes suggested last week that the UFC may be divided over the current crisis. Unusually for an important document, the circular letter was signed by the UFC's recently appointed secretary, Finlay Scott, rather than the chief executive, Sir Peter Swinnerton-Dyer. The UFC spokesman says Swinnerton-Dyer was attending a meeting in Scotland when the letter was sent out, and the priority was to inform vice-chancellors of the situation as quickly as possible. But the decision to suspend the bidding system had been taken at a UFC meeting nearly two weeks before.

In the long term, universities fear they may be forced to charge students fees for teaching in order to maintain quality in the face of expansion. A working group of the

Romanian drug trial halted

London

FOLLOWING advice from the World Health Organisation (WHO), the Romanian Health Ministry has halted trials of the controversial 'AIDS drug' FLV 23/A in children and adults in Bucharest infected with human immunodeficiency virus (see *Nature* 347, 606; 18 October 1990).

A WHO team visiting Romania last week concluded that the drug has "no likely therapeutic value", said David Heymann, an AIDS specialist at WHO.

The organizers of the trials had provided no detailed information on the drug's chemical structure and method of manufacture, Heymann says. Also, there appeared to have been no pharmacological studies in humans or animals, and no scientific evidence for antiviral action. WHO is now conducting its own laboratory tests to check for antiviral activity.

Peter Aldhous

EC ministers agree compromise target

London

EUROPEAN Community (EC) environment and energy ministers, meeting in Luxembourg on 29 October, agreed to stabilize EC emissions of carbon dioxide at current levels by 2000. The agreement allows EC ministers to present a superficially unified front at the World Climate Conference ministerial meeting in Geneva next week, but will be attacked by nations opposed to emissions controls as a vague statement, allowing some member states including Britain to stabilize emissions at a later date.

European Commission officials had hoped that stabilization by 2000 could be agreed, with only the 'less developed' member states of Spain, Portugal, Greece and Ireland allowed a more extended schedule.

Peter Aldhous

CVCP and the Committee of Directors of Polytechnics has considered future mechanisms of financing higher education (see *Nature* 347, 4; 6 September 1990), concluding that teaching costs may increasingly have to be borne by students. Professor John Ashworth, director of the London School of Economics and chairman of the working group, believes that such changes in the financing of higher education should be introduced centrally, by government, rather than be forced upon universities and polytechnics by financial attrition. But the Department of Education and Science says that decisions on fees are a matter for individual institutions.

Peter Aldhous

US on track for stabilization

Washington

TOTAL US greenhouse gas emissions will be at about the same level in 2000 as they were in 1987, even with no efforts at control beyond those already planned, a federal official said last week in Washington at a government conference on global change.

Alexander Cristofaro, director of the air and energy policy division of the Environmental Protection Agency (EPA), released results of a study showing that existing policies such as the regulations governing garbage landfills will have sufficient impact on production of greenhouse gases to counteract the expected increases in carbon-dioxide emissions by the end of the century.

The EPA study uses a "carbon equivalency" calculation (first developed by the Intergovernmental Panel on Climate Change [IPCC] for its report earlier this year) to compare the relative atmospheric impact of carbon dioxide, CFCs, methane and other greenhouse gases. Calculating the time these gases remain in the atmosphere, the EPA arrived at total 1987 US emissions equivalent to 2,328 million tonnes of carbon. By 2000, carbon-dioxide emissions will have gone up, but CFCs and methane will be down and other gases will remain relatively constant, thanks to regulations already in place. Total emissions at the turn of the century will be 2,332 tonnes — almost unchanged from the 1987 levels.

US officials hailed the figures as a triumph for the administration's "no regrets" environmental policy — the position that, in the absence of clear evidence that the greenhouse effect is upon us, the government should commit itself only to climate-change measures that have other benefits such as pollution control or energy efficiency.

Should the measured increase in global temperatures be due to something other than global warming, such measures — and their economic and social costs — will not have been in vain, presidential science adviser D. Allan Bromley argues in an article in this month's *Issues in Science and Technology*.

But environmentalists warn that the EPA study makes over-optimistic assumptions about the energy-efficiency incentives of the Clean Air Act and underplays the fact that emissions will rise considerably above 1987 levels during the 1990s before many of the policies have a chance to show an effect. And both critics and the EPA note that after 2000, total US emissions head for the stratosphere — reaching 2,430 million tonnes by 2010 — unless the United States adopts additional policy measures.

Nevertheless, the EPA report does

show that initial reductions can be achieved at minimal additional cost. And although the IPCC carbon-equivalents methodology may be in some ways flawed (it does not, for example, take into consideration sulphur dioxide, which is the principal industrial cause of acid rain, but may counter global warming by increasing the reflectivity of clouds), "it's probably the best we can do at the moment", says William Nitze, a former State Department official who is now president of the Alliance to Save Energy, "I don't think it's biased in any particular direction."

The most significant conclusion, he points out, is that President Bush could announce tomorrow that the United States would commit itself to achieving no more than 1987 total greenhouse gas levels by the year 2000, and it would cost him virtually nothing. Will Bush make such a declaration, perhaps in time for this week's World Climate Conference in Geneva? "Don't count on it", says one National Oceanic and Atmospheric Administration (NOAA) official. Without the backing of the other industrialized nations, he says, the United States is reluctant to make any promises

that it might someday regret if the scientific assumptions of greenhouse warming change.

Others have a different explanation. "The administration has always felt that target-setting is a slippery slope and they don't want to get started down it", says Alan Miller, of the University of Maryland's Centre for Global Climate Change.

Achieving 1987 emissions level in 1990 is encouraging, but even the most conservative administration officials recognize that more needs to be done if the greenhouse effect is real. Cumulative atmospheric totals of greenhouse gases are the quantities to worry about most, and continuing to pump an equivalent of over 2,000 million tonnes of carbon into the atmosphere each year will not reduce atmospheric concentrations, or slow global warming.

Indeed, few expect a strict no-regrets policy to be around for long, especially if scientific evidence continues to point to a climate on the verge of disaster. The NOAA official concedes that "all these current environmental measures fit into the no-regrets policy. But how many more of these things we can find if faced with global warming remains to be seen."

Christopher Anderson

Australia sets example

Melbourne

THE Australian government has agreed to stabilize at 1988 levels the emission of greenhouse gases by the year 2000 and to reduce them by 20 per cent by the year 2005, provided there is no damage to the Australian economy or to industry's international competitiveness. The agreement does not include the greenhouse gases, principally CFCs, that are already scheduled to be phased out under the Montreal Protocol and is thus considerably more ambitious than plans being put forward in the United States (see above).

The Minister for the Environment, Ros Kelly, will take Australia's commitment to the Second World Climate conference in Geneva on 6 November, stressing that Australia will act only if other countries adopt similar targets.

Australian industry is not confident that the targets can be met without considerable harm. According to David Nolan, director of the Confederation of Australian Industry, there may not be an overall effect on the economy, but individual industries will suffer from the 20 per cent cut. "It's impossible for the power industry, for instance, not be affected."

The Industry Commission will assess the cost of the 20 per cent reduction target to industry and will report to the government by the end of 1991. But this week,

the Department of Primary Industry and Energy announced a series of programmes designed to bring about immediate reductions in greenhouse gas emissions.

Focusing on the consumer, the measures aim to reduce Australia's annual energy bill by A\$1,500 million by 2005 and to reduce greenhouse gas emissions by 14,200 kilotonnes of carbon dioxide each year. Information kits on energy conservation will be distributed to each household; all government housing and commercial buildings will have energy managers and will be built using energy-efficient technology and all government departments and agencies will be required to report annually on their energy use and energy efficiency.

In 1986–87, according to George Wilkenfeld, an energy consultant to state and federal governments, Australia produced 279,000 kilotonnes of carbon dioxide, 6,400 kilotonnes of methane and 300 kilotonnes of nitrous oxide.

Wilkenfeld believes that the bulk of the 20 per cent reduction can be achieved only through the cooperation of industry. "Energy-efficient measures, used in industry, can cut carbon dioxide emissions by up to 30 per cent. The first 20 per cent reduction can actually be achieved at a profit because of the savings in power wastage."

Tania Ewing

Untapped data in vaults

Washington

ALTHOUGH it is not the sort of thing they always like to talk about, oil and gas companies spend millions on environmental measurements each year, collecting satellite, atmospheric, geological, and oceanographic data that could prove invaluable to climate change researchers. Invaluable, that is, if the scientists could only get their hands on it. Fear that the data might be misused has forced industry to dub most environmental data "proprietary", and guard it jealously.

The problem is "environmentalists", says Mark Settle, manager of integrated exploration research at the oil and gas company ARCO. "The concern is that we might be identified [as environmental offenders] by the release of that data." Because many of the measurements are collected so that oil and gas companies can catch environmental problems (and presumably fix them), they can contain evidence of gas leaks, pollution, and oil spills — not the sort of information a company would want to release unprompted. Some of the data, such as satellite observations and geological measurements, might also be useful to competitors looking for untapped resources.

Some cooperation between industry and researchers on previously proprietary data has already begun, largely through the efforts of the industry sponsored Gas Research Institute (GRI), an independent body set up (with funding from a percentage tax on gas transactions) by the federal government to investigate environmental and energy efficiency implications of gas use.

Working with the US Environmental Protection Agency (EPA), GRI has started a programme to measure leaks in natural gas pipelines. About one per cent of all methane produced is lost to the atmosphere in pipeline leaks, and perhaps

twice that lost elsewhere, says Kathleen Hogan, chief of the methane programme in EPA's office of air and radiation research. Although the total — around three to four terragrams — is less than the amount of methane released by landfills and livestock, it represents a greenhouse gas source that may be relatively easy to reduce.

But the methane leak study also points to a limitation in data collected by industry. Because gas companies are more interested in knowing whether or not a leak exists, than exactly how much methane is escaping, instrumentation is relatively coarse and rarely calibrated. Comparing readings from different sites and companies quickly proved impossible, Hogan says. As a result, EPA and GRI researchers will have to take their own measurements as the study continues over the next several years.

Other data may prove more useful. Satellite measurements commissioned by industry to spot potential oil and gas deposits have provided some of the best geological maps of large parts of the world. Similarly, the search for fossil fuels has led to a great deal of mapping efforts of the ocean floor and considerable research on oceanic chemistry. Nevertheless, "We have to be careful that the data is not splattered around the public out of context, because it might be damaging", says Howard Reiquam, director of GRI's environment and safety research department. In the past, GRI and the American Petroleum Institute have helped by collecting pollution data from industry and "homogenizing" or contouring it to hide the individual — and potentially self-incriminating — data points, says Reiquam. Establishing an industry clearinghouse for greenhouse data is an idea worth pursuing, but one that may still be some years off, he says.

Christopher Anderson

Carbon currency proposed to cut emissions

Washington

IN a world faced with global warming, carbon may become a virtual currency of international trade, at least if policy makers follow the recommendations of a new report* by the Washington-based Brookings Institute. After finding flaws in four common strategies (carbon taxes, across-the-board cuts, per capita targets and cuts based on historical emissions) to encourage global reductions in carbon emissions, the Brookings report proposes a new one: a scheme in which nations are issued "carbon chits" based on projections of populations and economic activity.

Countries may release only as much car-

bon as their chits allow; if they need more, they must buy them from other countries.

The advantage, according to the report, is that underdeveloped — but populous — nations such as India and China could sell their unused chits to carbon hogs such as the United States, a process that would both help to pay for energy-efficient technology in the developing nations, and also encourage the industrialized nations to cut their own emissions or risk going bankrupt.

Christopher Anderson

Controlling the Greenhouse Effect — Five Global Regimes Compared. The Brookings Institute, 1990.

Abuses in Sudan

London

As Western aid donors predicted a return to the famine conditions of the mid-1980s in Sudan last week, and accused the military government of obstructing their relief efforts, news also emerged of the extent of the abuses of academic freedom currently being perpetrated in the country. Since the elected government of prime minister Sadiq El-Mahdi was overthrown in a military coup in June 1989, the fundamentalist Islamic government, led by Lieutenant General Omar Hassan Al-Bashir, which replaced it has tried to reshape the educational system to fit with its ideals. Rakiya Omaar, from the human-rights pressure group Africa Watch, who has just completed a report on academic freedom in Sudan, says that the vice-chancellors of Sudan's four universities have been replaced by government appointees, and large numbers of academics have been dismissed.

Of particular concern is the fate of a number of academics, probably more than 20 in total, imprisoned without trial. One of these, Professor Moneim Attia, from the University of Khartoum, embarked on a hunger strike on 1 October, after being detained since January, beaten and tortured. Omaar believes that Attia has now discontinued his hunger strike. But Attia's academic colleagues in Britain still fear for his safety, given the appalling sanitary conditions of the Shalla prison in western Sudan in which he is held.

Salah Bander, a Sudanese geneticist now working at the Institute of Animal Physiology at Babraham near Cambridge, believes that Attia has been imprisoned because of his scientific work. In 1988, and with the support of the previous government, Attia set up an International Heat Stress Research Centre in Khartoum, winning finance from the Saudi Arabian government. Bander says that international scientific links are opposed by the governing revolutionary council, and adds that Attia's work on the problems of heat stroke among Sudanese industrial workers will also have been unpopular with a government committed to a rigid programme of industrialization. Although he is a former communist (not uncommon among Sudanese academics), Attia is not thought to have been politically active since last year's coup.

Professor John Bligh, a retired environmental physiologist formerly at Babraham, who is a trustee of Attia's research centre, hopes that pressure from the international scientific community can secure Attia's release. Professor Farouk Ibrahim, imprisoned apparently for teaching the theory of darwinian evolution at the University of Khartoum, was released earlier this year after international protest.

Peter Aldhous

Storm over funding policy

London

As part of a series of controversial cutbacks, the UK Medical Research Council (MRC) is to withdraw funds from an internationally renowned research unit in Cambridge specializing in molecular neurobiology, and is expected later this month to announce the closure of a blood-pressure research unit in Glasgow, the only one of its kind in Britain. The cutbacks mark the beginning of a tough new era for MRC funding, with a budget that is shrinking in real terms.

The cutbacks have been sharply criticized by many biomedical researchers, but are necessary, the MRC says, to generate cash for new research initiatives. With wage inflation running at 9 per cent and large sums earmarked for specific programmes such as AIDS research and the human genome initiative, the MRC is predicting a deficit in the coming year. This is rumoured to be as high as £6 million, although the MRC insists it is nearer to £1 million.

The work of each MRC research unit is reviewed every five years, and in keeping with the MRC's policy of backing individuals rather than research units *per se*, it is common for units to close when their directors retire. Given this perspective, the closure of the molecular neurobiology and blood-pressure units may not appear unusual, as their directors retire in 1992 and 1994 respectively. But this year, the MRC has introduced a new policy of assessing the funding of existing work in units scheduled for closure. The result has been the immediate shut-down of two of the six research groups within the molecular neurobiology unit, and the much-criticized decision to close another unit under review, the cryobiology unit in Cambridge, against the advice of expert referees (see *Nature* 346, 685; 23 August 1990).

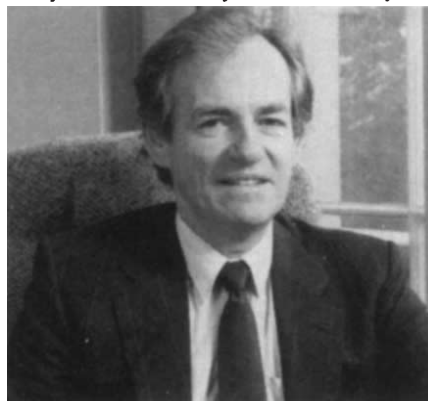
The early closure of the two research groups at the molecular neurobiology unit will save the MRC about £150,000 a year, but has dismayed unit members and other prominent neuroscientists who had assumed the MRC would stick to its normal practice of allowing units earmarked for closure to run down gradually through 'natural wastage'. Unit director Eric Barnard describes the decision as a "savage cut" — the two targeted groups contain much of the unit's vital expertise in recombinant DNA techniques, he says.

Nicholas Winterton, head of the MRC secretariat, defends the MRC's new policy of reviewing ongoing work, which he says allows a "much more focused search for savings" than simply considering new proposals. He says cuts in existing units are bound to attract more adverse publicity than refusals to fund new proposals,

but says the MRC is anxious to avoid all its resources becoming tied up in long-term commitments.

In response to criticism of the closure of the cryobiology unit, Dr Dai Rees, secretary of MRC, wrote to the directors of all MRC research units explaining the new policy. The MRC had been unable to fund some 400 alpha-rated project grant applications in 1989–90, he said, making it necessary to look for "what savings we can in ongoing MRC teams to release money for new work".

But Rees's explanation has not satisfied many staff affected by the cuts. Last year,



Dai Rees — savings are necessary "to release money for new work".

referees engaged in the routine review of the molecular neurobiology unit unanimously recommended that funding be continued. The decision to cut the two groups at the molecular neurobiology unit, despite their alpha ratings, was based on a recommendation from the MRC strategy committee, a subcommittee of MRC's council which examines funding priorities after peer review by the research boards. Critics say that the strategy committee has robbed the old research boards of much of their power and has failed to explain many of its recent high-level decisions or priorities. According to one research board member, Geoffrey Burnstock of University College London, the time has now come for the "relative roles of the traditional boards and the strategy committee to be clarified".

The MRC also appears to be adopting a tougher stance on public debate of its funding decisions, with one member of the molecular neurobiology unit under threat of disciplinary action over a letter published in *Nature*. In September, Rees told Barnard to warn his staff against engaging in any "improper criticisms of Council's policies or decisions" that may draw Rees into the public arena. Nevertheless, a letter to *Nature* by Rees himself (see *Nature* 347, 116; 6 September 1990) prompted the leader of one of the axed research groups, Penny Barnard, wife of the unit director, to respond (*Nature* 347,

Museum with friends

London

An organization of 'Friends' of the Natural History Museum got off to a shaky start last Thursday amid continuing animosity between the museum's director, Neil Chalmers, and the trade union representing scientists at the museum.

The embryonic body is the brainchild of Henry Barlow, an amateur entomologist based in Malaysia. The chairman of the meeting was Jonathan Porritt, former chairman of the environmental pressure group Friends of the Earth.

Chalmers emphasized that the museum is now raising a growing proportion of its own funds, and said that £9 million out of a total budget of £34 million is generated internally. Some of this is income from the museum's development fund, whose first appeal is now just £400,000 short of its £5 million target.

Chalmers sees one function of the museum's Friends as that of providing support and argument in the "battle with the government that goes on year by year". But he also noted last week that "the first thing about friends is that they are friendly".

The appointment of an interim steering committee evoked some contrary opinions. Although Chalmers had nominated two staff members (John Peake, associate director, and Jill Nadolski, co-ordinator of the museum's membership scheme), he opposed the nomination of William Lindsey, a palaeontologist who is a member of the National Executive of the Institution of Professionals, Managers and Specialists (IPMS), which is in dispute with the museum over the effects of its corporate plan on members of staff.

But the tension at the museum appears to be abating. A third one-day strike at the Natural History Museum called for 31 October has been narrowly averted. Most of the cuts in research posts outlined in the corporate plan have now been made, and the number of enforced redundancies is likely to be fewer than ten. Henry Gee

324; 27 September 1990), calling on the MRC to explain the financial position underlying the need to make such drastic cuts in an established unit at short notice.

In a confidential letter to Eric Barnard, Rees described this letter as "sufficiently misleading as to be damaging to the Council", and insisted that action should be taken over what he termed a "serious breach of discipline". Barnard has refused to discipline his wife, and is waiting for Rees's response. Winterton says that a clause in MRC staff contracts forbids public comment against MRC policy, but maintains that the MRC is "a good deal more relaxed" in its attitude to public dissent than most employers.

David Concar & Peter Aldhous

Industrial funds flow freely

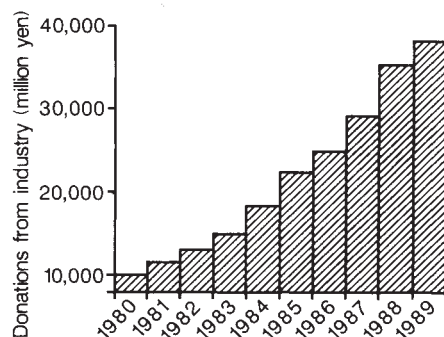
Tokyo

ALTHOUGH no one seems to have noticed, the latest budget figures from the Ministry of Education, Science and Culture (MESC) reveal that an astonishing change has taken place in Japan's universities.

Despite Japanese academics' vociferous insistence on their independence from Japan's industrial goals, the figures show that, in the coming year, funds provided by industry for research at universities will amount to ¥57,650 million (\$461 million), a sum almost equal to the ¥million MESC plans to spend on research grants.

Almost 80 per cent of the money comes in the form of 'donations', which now total about ¥40,000 million (\$320 million) a year after a decade in which they have quadrupled. Growth is still at 20 per cent a year, far above the rate of inflation and the rate of increase of government funds. Most donations are in the form of private research grants for faculty members, according to Wataru Soga of the research cooperation division of MESC, but a small

part is also used to establish chairs and new research divisions at the universities. The new Research Centre for Advanced Science and Technology at Tokyo University provides one example with eight research divisions funded by companies such as NEC, Nippon Telegraph and Telephone (NTT), Nippon Steel Corporation and Ricoh. Ten out of 24 of such new research divisions in Japan are headed by foreigners, employed on short-term (two-



to five-year) contracts.

Although industrial funds may be large, they are not evenly distributed. Members of the faculty of science at Tokyo University say that, apart from their colleagues in computer science, the faculty benefits little from donations. Most of the money goes to the faculty of engineering and the faculty of medicine through connections with electronics and pharmaceutical companies. Personal relationships between powerful faculty members and companies are critical in determining who gets donations and a steady supply of bright young researchers is a reward for the donor company.

MESC has also put government money into schemes intended to encourage joint research between universities and industry. This year the programme contains 730 projects and costs ¥3,755 million (\$30 million). More than 80 per cent of the money comes from industry and the remainder from MESC. The universities provide the research facilities and the researchers. Industry is also able to carry out "commissioned" (*jutaku*) research, at universities in which it provides young researchers for specific projects. Industry provides all the funds and will spend ¥7,110 million (\$57 million) this year.

Despite the growth of industrial funds, much red tape remains. To obtain a small grant of ¥1 million (\$8,000) from a company, Shahid Siddiqui of the Laboratory of Molecular Biology of Toyohashi University of Technology first had to apply to the company for the grant and then "donate" it to the university.

Only after it was "officially" recognized by the university and MESC could he actually use the money for research, six months later.

David Swinbanks

Still over the horizon

Washington

IN spite of the increased availability of more effective methods of predicting or detecting genetic disease, a survey of US companies shows little change in the use of genetic testing in the workplace over the past seven years. More interestingly, the report*, released last week from the US Office of Technology Assessment (OTA) finds that fewer companies anticipate the future use of genetic testing than did in a similar earlier survey conducted in 1982.

The OTA report surveyed 500 of the largest US corporations, the 50 largest utility companies and 33 major unions, in order to provide comparability with the 1982 survey. In addition, the scope of the survey was broadened to include smaller companies.

In the questionnaire, a distinction was made between genetic screening, which focuses on an individual's pre-existing genetic make-up and generally involves a one-time test to detect a single trait, and genetic monitoring, which requires multiple tests over time to determine the exposure of an individual to hazardous materials in the workplace.

OTA found that of the 330 (62.4 per cent) largest US companies that responded to the 1989 survey, 12 reported current use of genetic monitoring or screening programmes, and eight reported past use.

Similar results were found in the 1982 survey, where of the 366 (65.2 per cent) of companies that responded, six had active screening programmes and 12 reported past use. Of the companies with active programmes, testing involved cytogenetic monitoring and biochemical genetic screening. No companies reported the use of direct-DNA monitoring or screening.

The availability of these new technologies poses a real dilemma. While genetic monitoring and screening could yield useful long-range, predictive information about occupational health risks, the rights and responsibilities of job applicants, employees and employers are not well defined. Only a limited body of law deals specifically with genetic monitoring and screening in the workplace.

In a recent initiative, Representative John Conyers (Democrat, Michigan) introduced a bill, entitled the Human Genome Privacy Act, which would prevent the disclosure of genetic information by government agencies, contractors or grant recipients to third parties without the individual's consent (see *Nature* 347, 221; 1990). Although the bill does not encompass private industry, the bill's supporters believe that government should take the lead and set a standard by which companies should respond. Diane Gershon

*Genetic Monitoring and Screening in the Workplace. OTA-BA-455 (Washington, DC: US Government Printing Office, October 1990).

Scattered largesse

Tokyo

A separate initiative by MESC to stimulate collaboration between industry and academic institutions is the 'regional joint research centre' scheme which is aimed at smaller, local universities in the remoter regions of Japan. Eighteen centres are now in place at universities in Kyushu, Hokkaido, and the main island of Japan. Five more will open next year.

The Cooperative Research Center of Kumamoto University in Japan's southern island of Kyushu, which was completed in 1988 at a cost of ¥500 million (\$4 million), is typical of the scheme.

Located in the Kumamoto 'technopolis', an industrial research park being promoted by the local prefectural government and the Ministry of International Trade and Industry (MITI), the centre has four permanent faculty members from Kumamoto University who are in charge of about 20 projects being carried out by graduate students and researchers from companies. The projects each have annual budgets of a few million yen and include such things as wind tunnel tests of skirts for hovercraft, investigation of chemical vapour deposition techniques to make high-temperature superconductors, and development of new high-speed image-processing systems.

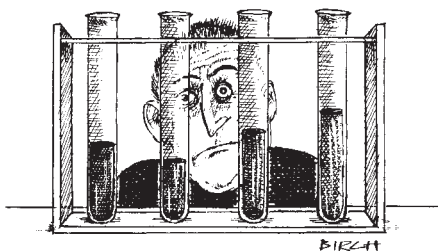
It is too early to tell how successful the centre will be. But a recent tour revealed that the centre's small team of researchers have plenty of space and equipment.

David Swinbanks

Chained to the bench

Washington

DEPENDING on one's perspective, one could say that Syed Zaki Salahuddin, a scientist working on AIDS, who had pleaded guilty to misdirecting federal research funds, received an easy sentence



last week. Or if one takes a different view of life at the laboratory bench, one might say the judge threw the book at him. Salahuddin, until recently a member of the National Institutes of Health (NIH) AIDS laboratory of Robert Gallo, was sentenced to spend what amounts to every Saturday for the next four years doing research — without pay.

Now a visiting researcher at the University of Southern California, Salahuddin will spend 1,750 hours investigating the viral origins of chronic fatigue syndrome as part of the sentence. US District Judge John Hargrove also fined him \$12,000, approximately the value of the NIH contracts he steered to Pan Data Systems, a Maryland-based contractor partially owned by his wife (see *Nature* 345, 99; 10 May 1990).

"I thought it would be a good idea to take advantage of his research abilities", says prosecutor Dale Kelberman. In some previous cases involving researchers, doubts were raised on their scientific integrity and they were sentenced to lesser tasks such as computer operating and data entry.

Christopher Anderson

EMBO

Fellowships for East

Munich

THE European Molecular Biology Organization (EMBO) will next year begin awarding new short-term fellowships to applicants from Eastern Europe who would like to work in a Western European laboratory. The programme will begin with ten three-to-six-month fellowships paid for from the *EMBO Journal's* profits.

According to John Tooze, executive secretary of EMBO, demand from science-starved Eastern Europe is likely to exceed by far the supply of funds available. He sees the new programme as a "gesture" that he hopes will "prime the pump" to induce other foundations or EMBO member states and Eastern European governments to expand the programme.

Steven Dickman

Germany turns clock back

Munich

THE German Bundestag last week passed a law making research on human embryos a criminal offence punishable by up to five years in prison and imposed severe restrictions on the use of *in vitro* fertilization (IVF). The law, which also includes limits on changes in the human germ line and other invasive genetic procedures, gives Germany the world's strictest embryo research regulations. The new law takes effect on 1 January 1991.

The ban sets a potentially dangerous precedent for research in Germany by calling into question the constitutional protection of freedom of research. But efforts by research organizations to incorporate exceptions to the ban were rebuffed by the shared view of all political parties that such research is unethical (see *Nature* 340, 254; 1989).

The legislation imposing the ban specifies that embryos may not be created for any purpose other than implantation in a woman's uterus. The number of egg cells that may be fertilized for any one IVF operation is limited to three. The use of embryos for any other purpose including research is to be considered a criminal act. Other countries, including the United Kingdom, permit research on embryos until they are 14 days old.

Significantly, although the German law forbids research involving the removal of so-called 'totipotent' cells (which could be used to create a new embryo if removed at the 8-cell stage of the dividing embryo or earlier), it will allow prenatal diagnostic testing of later stages. Physicians can test for genetic diseases on cells removed from the 32-cell or 64-cell stage, known as the morula, without harming the embryo. If a disease is detected, the embryo can be discarded before implantation.

According to Henning Beier, a professor of anatomy and reproductive medicine at the Technical University in Aachen, the law does not present an "insurmountable barrier" to embryology and diagnostics research. Beier says that almost all embryology research in Germany is carried out using animals.

Two large German research organizations, the Max Planck Gesellschaft and the Deutsche Forschungsgemeinschaft, declared in 1988 that they would observe a

Correction: RITE figures wrong

In the article entitled 'Japan sees gold in warming' (*Nature* 347, 703; 25 October 1990), all budget figures were inadvertently given ten times too large. The money contributed by Japanese companies to the Research Institute for Innovative Technology for the Earth (RITE) is ¥5,000 million (\$40 million) not ¥50,000 million (\$400 million). □

self-imposed moratorium on embryo research until there was a law (see *Nature* 333, 791; 30 June 1988).

Both research organizations supported sections of the law that ban the cloning of human beings and the creation of human-animal chimaeras. But some researchers wanted to allow exceptions to a ban on manipulating the germ line in cases where gene therapy might be of benefit. Harald zur Hausen, director of the German Cancer Research Centre in Heidelberg and an outspoken opponent of the law, said that pressure would soon grow to revise the law in order to allow gene therapy.

The new law would have banned completely the use of preimplantation diagnostic tests to select the sex of the embryo to be implanted. But at the last minute the government chose to add an exception allowing selection against embryos carrying severe sex-linked disorders" such as Duchenne muscular dystrophy in which case the son of a mother who is a carrier has a 50 per cent chance of developing the crippling disease, whereas daughters are rarely affected. Therefore male embryos are not usually implanted.

The opposition Social Democrats and Green Party voted against the law because it did not go far enough. The law gives "legislative blessing to eugenic measures" in Germany [for the first time since 1945], says Anna Waldschmidt, a spokeswoman for the Green Party.

Green parliamentarian Marie-Luise Schmidt strongly criticized the government during a Bundestag debate on 24 October for allowing embryos carrying genetic diseases to be discarded.

Schmidt said that this provision comes dangerously close to the selective measures included in the Nazi laws for the prevention of genetically diseased offspring. "People who were born with muscular dystrophy are apparently meant to find themselves in this law as the objects of a . . . negative selection that defines them as worthy of extermination", she said.

The Green Party opposes IVF in general as a "massive experiment on human subjects," because the techniques do not allow for a high rate of success. Instead, the party urges support for more research into the causes of infertility and into solutions other than IVF.

The Social Democrats opposed the law because the government had failed to address questions they had raised about artificial insemination and inheritance law.

Abortions, which are permitted under different circumstances in what were once East and West Germany, are not affected by the new law.

Steven Dickman

IPCC's climate change mindset

SIR—A careful reading of the recent reports of the Intergovernmental Panel on Climate Change (IPCC) suggests that continuing uncertainties about climate¹ may be less of a threat to rational forward planning than unwarranted certainties, or mindsets, of several years' standing. Moreover, the procedures being used for communicating analysis and policy proposals about climate change make it difficult to identify greenhouse mindsets among the handful of real facts on which is balanced the inverted pyramid of a huge mass of policy proposals².

IPCC, set up by the World Meteorological Organization and the United Nations Environmental Programme, formed three international working groups (WGs) in November 1988 to make a scientific assessment of the problem (WG1), the potential impact of climate change (WG2) and to suggest strategies for response (WG3). Each group formed about half a dozen subgroups, each with 6–12 expert contributors. There were arrangements for editorial coordination; indeed, the project was a huge writing and editorial task. After many meetings and fax exchanges, group reports were completed in June this year^{3,4}. IPCC continues.

Three indices of climate change — temperature, rainfall and sea level — attract most attention. WG1, on its 'business-as-usual' assumption, says that the likely increase of temperature by 2025 will be about 1°C, which is at the low end of predictions during the period 1987–88 (ref. 6). But WG1 acknowledges "low confidence" in its estimates of regional rainfall change, especially in regard to "timing, magnitude and regional patterns". Similarly, the WG1 estimates of sea-level change (increases of 20 cm by 2030 and 65 cm by 2100) are also much smaller than had previously been thought possible⁶.

While the two-year WG1 study properly reduced previous estimates of predicted climate change and prudently emphasized the uncertainties, WG2 is innocent of corresponding amendments. Indeed, WG2 says that, "in an ideal world", it would have assessed the impact of climate change *after* WG1 had completed its study, but that the need for the work to proceed in parallel precluded that. Instead, WG2's report is based on estimates of climate change generally current in 1988.

But it is not now possible to identify the inaccuracies in the WG2 report that derive from its unsustained assumptions. This is especially the case because WG2's 'Policymakers' Summary' is largely qualitative (except for numbers for sea-level rises). The lack of quantitative data is sometimes the consequence of uncertainty, as in the

statement that it is unclear whether global agricultural potential will increase or decrease. But most potential impacts are presented more confidently, as BLACK or WHITE, but with the frequent use of the words "may" and "could". Overall, the lack of quantitative data in the WG2 report makes it impossible to make scale corrections of the impacts in the light of the new WG1 findings.

Where quantitative data do appear in the WG2 report — on sea level, in particular — the use made of them reveals a mindset towards outdated predictions, and especially towards a sea-level increase of 100 cm, now forecast by WG1 for 2130. WG2's first recommendation on coastal zones is, nevertheless, that there should be an effort to assess risks caused by a rise of 0.3 to 0.5 m, implicitly acknowledging that this has not been done.

Although WG2 at first¹ used three scenarios — increases of 0.5 m, 1.0 m and 2.0 m — the final report⁴ does not distinguish between the impacts of the three alternatives although (in Chapter 6) rises of 1.0 m are mentioned 27 times, rises of 0.5 m or so only 6 times. Clearer evidence of the mindset is apparent in the solitary table (6.1), in which the costs of coastal protection for 50 countries are calculated only for a sea-level increase of 1.0 m.

Inconsistencies of sea-level scenarios persist in the June 1990 version of the WG2 report, in phrases such as "nevertheless the impacts based on 1–2 metres . . .". But the final version of the report (in section 5.1.2) states, "A rise of only 25 cm or more . . . could render uninhabitable nations such as the Maldives . . ." and, elsewhere, "the existence of entire countries such as the Maldives . . . [is] imperilled by a rise of only [sic] a few metres". In WG3, the Maldives are mentioned (among others) as nations within three metres of present sea level.

Perhaps the most striking illustration of the greenhouse mindset is to be found in the 'Policymakers' Summary' of WG3, where (in section 7.1) the vital discussion of coastal zone management begins with the statement that, on the assumption of continued high emissions, there will be an increase of "mean sea level [of] 65 cm (with an uncertainty range of 30 to 100 cm) by the year 2100", and then continues, "if sea level rises by 1 metre, hundreds of thousands of square kilometres . . . could be inundated". The remaining discussion deals only with the consequences of rises of 1 metre even though the new findings of WG1 postpone that event well into the century after next.

Drafting difficulties are understandable in international committee deliberations, but the persisting inconsistencies are also valuable indicators of the mindset of many

contributors. One must, of course, sympathize with those who have invested time in calculating the impact of events for which the predictions are then changed, but the conclusion must be that the findings of WG2 and WG3, as reported in June this year, are fatally flawed in three respects: (1) IPCC's methodology; (2) the credibility of the 1-m rise of sea level; and (3) the absence of numerical estimates in the potential impacts described by WG2.

Uncertainty is not the chief concern, for in principle that could be removed by further research. But the IPCC reports show that this has not yet been done. The more serious worry is the ease with which the prose text of documents such as these can be divorced from fact and real-world numbers by the ubiquitous word-processor, with the danger that particular conclusions about greenhouse impacts will retain their currency and force even when the assumptions on which they are based have been changed or rendered irrelevant.

This in turn raises the concern that the greenhouse problem is not merely an inverted pyramid of knowledge based on a handful of facts², but that the facts may now be buried in a pyramid of much-manipulated reports. Indeed, the documents circulated in June 1990 may represent the last identifiable connection between the supposed greenhouse impacts and the facts. It may, nevertheless, be some comfort that if the assessment of impacts and the devising of response strategies is to continue to be independent of more refined research, further research may not be necessary.

BRIAN J. O'BRIEN

12 Caithness Road,
Floreat Park, Western Australia 6014

1. *Nature* **347**, 1 (1990).
2. White, R.M., Annual General Meeting of US Academy of Sciences, Washington DC, 24 April 1989.
3. *Scientific Assessment of Climatic Change*, Report from WG1 to IPCC, June 1990.
4. *Potential Impacts of Climatic Change*, Report from WGII to IPCC, June 1990.
5. *Formulation of Response Strategies*, Report from WGIII to IPCC, June 1990.
6. *The Changing Atmosphere: Implications for Global Security*, Toronto, June 1988.
7. Report of the Second Session of IPCC WGII, Paragraphs 4.2.1 and 4.7.2, Geneva (October 1989).

Unicycling

SIR—Thanks to John Maddox for giving an update on the kinematics of bicycling (*Nature* **346**, 407; 1990). I believe the unicycle should be the optimum general exercise device. Not only does it require vigorous leg exercise, but it forces the user rhythmically to employ muscles of the abdomen, diaphragm, back and arms. Most importantly for those of us nearing tottering old age, the unicycle requires constant adjustment of balance. Anything new on the kinematics of unicycling?

JOHN A. JOHNSON

University of Nebraska Medical Center,
Omaha, Nebraska 68134, USA

Reply from the Gulf

SIR—The leading article "Beyond the trouble in the Gulf" (*Nature* 346, 683; 1990) theorizes about the alleged failings of economic and social policies in the Gulf countries on a basis of misunderstanding and even unfamiliarity with the facts.

The article offers three arguments why peace will not come to the Middle East even after the current crisis is over. First, it says the Gulf countries have squandered their oil wealth and failed to share it with less fortunate countries. "The bulk of what other oil-producing states earn [other than Iraq] is routinely sent back to the purchasers of oil, usually in the form of credits controlled by named individuals".

To examine these arguments, let us take the example of Saudi Arabia. No other country has shared its wealth so generously with the poor, developing countries.

Safety first

SIR—The dilemma posed by issues such as the safety and economics of bovine somatotropin (bovine growth hormone) (*Nature* 347, 11; 1990) was best expressed by that master of the two-liner, Alexander Pope, a hundred years before the luddite-like solution ascribed by Lesser to opponents of the hormone:

In words, as fashions, the same rule will hold,

Alike fantastic if too new or old:

*Be not the first on whom the new are tried,
Nor yet the last to lay the old aside.*

— Alexander Pope,
Essay on Criticism,
Part II, line 133 [1711]

Physicians have told me of a medical aphorism based evidently on the Pope dictum, or perhaps on harsh reality: don't be the first to try a new drug or the last to give it up.

BURTON L. APPLETON

PO Box 16057,
Alexandria,
Virginia 22302, USA

Noted scientists

SIR—I should like to add to the list of scientific figures to have appeared on banknotes (*Nature* 347, 415; 1990) Erwin Schrödinger on the Austrian 1,000-schilling note, and — although not on a par with the other scientists — Sigmund Freud on the 50-schilling note and Eugen Böhm-Bawerk (1851–1914), an Austrian political economist, co-founder of the marginal utility school, on the 100-schilling note. All three notes are currently in circulation.

FRIEDRICH KATSCHER

Mariahilfer Str. 133,
A-1150 Vienna, Austria

Saudi Arabia provided aid worth US\$59,700 million between 1973 and 1989. This figure works out to 5.45 per cent of gross national product (GNP) which is by far the highest ratio of aid to GNP in the world. Out of this figure, fully 57 per cent or \$34,380 million was given in the form of grants. More than 70 developing countries have benefited from Saudi aid, of which 38 are in Africa, 25 in Asia and 7 in other developing countries. Saudi Arabia is second only to the United States in terms of the total value of aid given in the period 1973–89. So much for the argument that Saudi Arabia, among others, has not shared its oil-wealth with less fortunate countries.

The second argument is even more absurd. It claims that these countries have neglected the development of social infrastructure, specifically education. "The oil-producers of the Middle East seem not to have turned their minds to creating a first-rate university somewhere." Again, the author has absolutely no idea of the tremendous progress made in Saudi Arabia in education and the social sectors over the past 15 years. The King Fahd University of Petroleum and Minerals in Dhahran is universally recognized as among the best of its kind in the world. It is not for nothing that its Saudi graduates are progressively taking over the oil sector in Saudi Arabia. But more than this, 2.8 million male and female students were enrolled in various educational institutions in 1990 compared with 573,000 students in 1970. The number of schools went up from 2,949 in 1970 to 16,476 in 1990.

In higher education, 122,000 male and female students were enrolled in 78 institutions of higher learning in 1990 compared with a figure of 6,942 students in 1970. The total number of teachers in higher education was 9,950 in 1988, which gives a student-teacher ratio of 11.5, surely among the best in the world.

Finally, the third argument truly reflects a lack of knowledge of the region. The authors says: "There has been too little investment in industrial enterprise, even in the downstream processing of oil itself". The truth is exactly the opposite.

Again, to take the example of Saudi Arabia, the entire economic philosophy of Saudi planning is to diversify the economy away from the production and export of crude oil alone. The successful breaking away from the 'monoculture' of oil can be seen from the fact that in 1975, crude oil and natural gas contributed 75 per cent to GDP, while it contributed only 23 per cent to GDP in 1987. This is a remarkable achievement indeed. And it is the result of using the oil wealth to enlarge the industrial base of the country.

Establishing advanced industrial cities

in Saudi Arabia started in 1970, and much of it already developed indicates a substantial commitment towards industrial development. Jubail and Yanbu were created from scratch in what is commonly recognized as one of the greatest planning, engineering and construction projects of all time. Oil refineries and petrochemical industries are today fully operational there. The petrochemical industries, most of which are already being expanded, command roughly 5–7 per cent of the world's installed capacity in bulk petrochemicals.

Hand-in-hand with the process of domestic industrialization, Saudi Arabia has successfully embarked on a course of downstream integration in the global oil market. The joint venture with Texaco, called Star Enterprise, is already a model of cooperation and integration. It commands roughly 11,000 retail outlets and three refineries, and assures the company of up to 600,000 barrels a day of Saudi crude. This programme of downstream is bound to grow in the future in the interests of pure commercial viability as well as the reciprocal security of supply and demand in the oil world. Besides, the Saudi economic Offset Program established in the mid-1980s requires the successful bidders for the sale of defence equipment to invest in partnership with the Saudi investors an amount equal to 25–30 per cent of the total programme contract. It was further required that the investment be in high-tech industries.

More than all of this, Saudi Arabia has succeeded in creating what is perhaps the ultimate welfare state in the world. State-of-the-art physical infrastructure is already in place. Health-care and education are free and widely available. Housing is subsidized and most Saudis own their own houses. Electricity, water and telecommunications are all abundantly supplied to meet the needs of the country.

HUSSEIN A. SEJINI

Ministry of Planning,
Riyadh, Saudi Arabia

SIR—The leading article "Beyond the troubles in the Gulf" (*Nature* 346, 683; 1990) well describes the causes of the current trouble in the Middle East. The unequal distribution of oil revenues, the lack of constructive development in the oil-producing states, and the corruption of the ruling governments have definitely contributed to the current conflict, but another major cause seems to be overlooked: the Palestinian-Israeli conflict. I cannot imagine a peaceful Middle East for as long as Israel continues to occupy Palestinian, Syrian and Lebanese land and as long as the five million Palestinians remain living in exile.

M. D. ABOUZAHA

PO Box 1087,
APO San Francisco 96555, USA

What will happen to science policy?

Harvard's celebration last week of a retired professor's birthday is a reminder that US academics have contributed powerfully to public policy.

Cambridge, Massachusetts

IN the end, we never learned where Harvey Brooks stands on the prospect of global warming, but nobody doubts that he regards it as deserving of the serious attention he customarily pays to public issues of this kind — and that he will write about it very soon.

Brooks is the now-retired engineering professor at Harvard who has become a legend in the United States and elsewhere for his indefatigable concern with technical issues in public policy — and for his habit of taking detailed notes of all the meetings through which he sits, including that with which Harvard's Kennedy School of Government celebrated his 75th birthday at the weekend. The eventual fate of these documents, apparently well-preserved, is a matter of concern, and perhaps sometimes of anxiety, to his erstwhile fellow-committee members.

Harvard had gathered for last week's occasion more than a hundred of these, including two former science advisers to the US President (Donald Hornig and Guyford Stever), several past and present members of the National Science Board and a generous sprinkling of past members of the President's Science Advisory Committee (PSAC) from its heroic pre-Nixon days.

Brooks's claim on the attention of academic colleagues and government officials alike is, by common consent, remarkable. A necessary but not sufficient explanation is a precociously solid reputation as a researcher (in the field of the mobility of atomic vacancies in metals and the nature of the inter-grain matrix of polycrystalline materials). But most of the others are personal.

Even at seventy-five, Brooks retains a compelling boyishness, both in appearance and in his readiness to be surprised and stimulated by notions of which he has not previously heard.

Linked with that is a beguiling artless innocence: all interesting questions deserve a serious answer, which is a matter of carrying out whatever calculations may be necessary, of rehearsing the arguments on one side and the other and of providing a literate analysis of their relative weight. In the world of the sound-bite in which emphatic opinions tend to carry the day, and in a field in which language is perpetually at risk of corruption by sociologese, Brooks stands out as a scholar who can write clearly.

But how, having come to a sound opinion of some issue of public policy, does one ensure that those who matter listen? Brooks's way is to join — or, as likely, to form — the right committee. One of his innovations of this kind was the Committee on Science and Public Policy, formed within the US National Academy of Sciences in the 1960s, which was the foundation of that organization's influence on the generality of technical public issues, not simply those on which the US Congress has commissioned a study.

His influence in that role, copiously illustrated by anecdote last week, seems to spring from his tenacity. He is the committee member who can always be counted on to have read the papers for the current meeting — as well as the previous minutes. The same diligence appears to stamp his function as an expert reviewer of about-to-be public documents: John H. Gibbons, director of the US Congressional Office of Technology Assessment, thanked him publicly at the meeting for 40 kbytes of comment on a document that has not yet seen the light of day. For what it is worth, Brooks was one of the midwives, nearly two decades ago, of Gibbons's own organization.

Although Brooks's reputation rests largely on his studies of domestic US policy in the fields of energy, environment and science policy, his influence overseas has been considerable. In the 1950s and early 1960s, for example, he was a regular member of the expert committees to which the Organization for Economic Cooperation and Development (OECD), then the Organization for European Cooperation and Development, entrusted the examination of science policy in particular countries. He argued the case, in the United States, in the early 1970s, for the creation of the International Institute for Applied Systems Analysis (at Vienna) and is now a regular member of the UN advisory committee on science and technology. Among other things, a willingness to catch an aircraft flight to almost anywhere seems to be an essential qualification for these pursuits.

So how does Harvard celebrate a career like this? First, with a more than decent dinner, at which Dr Derek Bok, the president of the university, raised the question whether it would be possible to replace Brooks's generation of policy advisors with new recruits of necessity innocent of the turmoil of the Second

World War. Then, with the traditional symposium devised to cover the range of the birthday boy's interests at which, last week, the most arresting contribution was that of Dr Thomas Schelling, the iconoclast economist who is best known for his contributions (while at Harvard) to strategic studies, but who now (as professor of economics at the University of Maryland), is powerfully advocating a gloomy view of the prospects for a treaty on global warming.

Acknowledging that there is a "vast amount of enthusiasm" for some kind of international action in the field, Schelling says he is "profoundly sceptical" that anything will happen. And "that makes me both a pessimist and an optimist".

The pessimism is easily explained. Schelling believes that enthusiasts for a treaty on global warming have underestimated the difficulties, not least those of persuading democratically elected governments to shoulder the costs of abatement, even if these are as little as one or two per cent of the annual gross national product for the rest of time. And in any case, if there are sums of money of that kind to be spent in the interests of the global condition, who will not sympathize with the views of states such as Bangladesh that they would prefer to have the resources now, as cash for the support of development of some kind, than as an investment in the abatement of emissions and the uncertain benefits thereof?

Schelling's optimism rests on two foundations — the argument that the predictions of the computer models of climate are, for the time being, "not very clever" and, more cogently, that for countries such as the United States, where agriculture accounts for some 3 per cent of gross national product, the costs of adaptation to climate change, while still uncertain, cannot be very large. Developing countries are another matter.

Challenged, Schelling says he would invest a small proportion of gross national product in the abatement of emissions, but that — given the choice, which is unattainable — he would prefer to see the same resources spent on the containment of population growth. But, in any case, he sees no way in which "countries such as China and India" will accept constraints on their emissions "unless we pay them".

Will reasonable Harvey Brooks be content with that, or will he take Schelling's opinions as a challenge? **John Maddox**

Black hole or no black hole?

E. S. Phinney

ASTRONOMERS are among the most international of scientists. Yet chauvinism surfaces on a cosmic scale where the anxious search for a massive black hole at the centre of our Galaxy is concerned. The prodigious amounts of energy emitted from tiny regions¹ at the centres of some nearby galaxies imply that they contain accreting black holes of masses ranging from 10^5 to 10^8 times the mass of our Sun ($M_\odot$). The orbits of stars in the unspectacular centre of our nearest sister galaxy (Messier 31, visible to the naked eye in Andromeda) indicate² that it contains at least $10^7 M_\odot$ in a form so dark³ and concentrated as to suggest the presence of a black hole⁴ of about $10^6 M_\odot$. Surely, some astronomers argue, nature cannot have been so unkind as to deprive our own magnificent Milky Way of such an ornament? On page 45 of this issue⁵, Yusef-Zadeh, Morris and Ekers report the latest result of the scrutiny of the centre of the Milky Way.

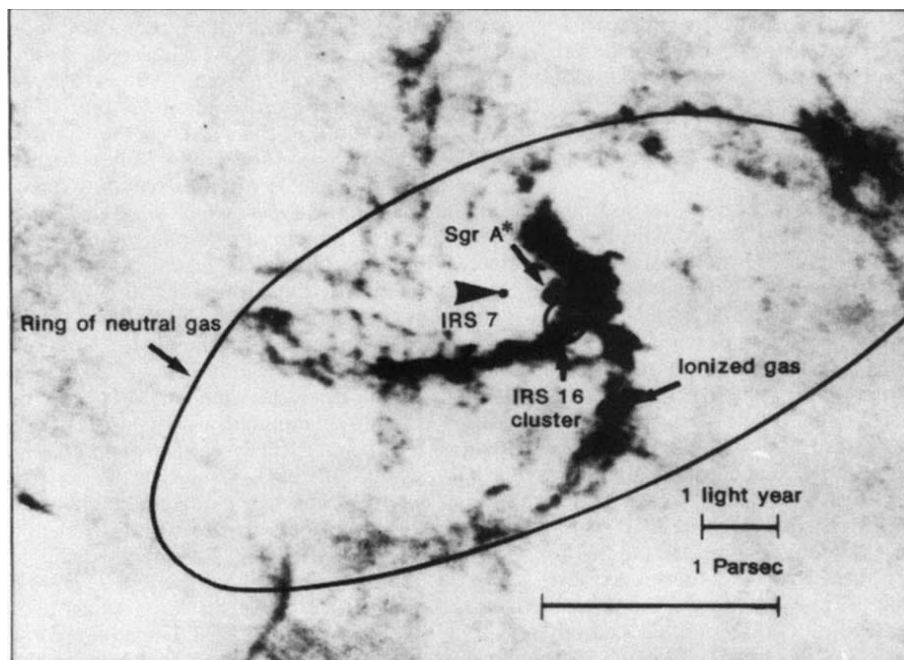
Dust and gas absorb most electromagnetic radiation from the centre. Only γ -rays and wavelengths between radio and infrared can reach us. Each of these shows an interesting, but confusingly different picture of the centre of our Galaxy⁶. Infrared observations reveal a dense cluster of several million stars within the central parsec (3.3 light years). The 100-km-s^{-1} velocities of these stars mean that there must be a mass of about $4 \times 10^6 M_\odot$ in the same region⁷. Much of this mass is in the observed stars. Subtle arguments intimate that part of this mass may be more concentrated than the starlight, but this is hardly irrefutable proof of the presence of a single supermassive black hole.

Accretion onto supermassive black holes in other galaxies reveals itself by ultraviolet radiation from the accretion disk, by high-speed (over $1,000\text{ km s}^{-1}$) flows of ionized gas, and by X- and γ -radiation (accompanied, according to theory, by the copious production of electron-positron pairs, although these have yet to be observed). Some also have powerful synchrotron radiation from plasma moving in jets at relativistic speeds.

Our own Galactic Centre displays feeble renderings of many of these characteristics. These have been adduced as evidence for a central black hole. Unfortunately, recent observations point to more prosaic explanations of most of these phenomena. The ultraviolet radiation and the ionized high-speed wind appear to originate from a group of very hot young stars⁸. The unusual radio source Sagittarius A* (the unique character and immobility⁹ of which make it the best

candidate for the signal flare of a massive black hole) has no coincident infrared source, and is clearly displaced from the origin of the wind and ultraviolet flux. A large flux in the 511 keV γ -ray line pro-

and Ekers⁵ of several faint radio sources clustered only a few light weeks away from Sgr A* may eventually help reveal its nature. Were Sgr A* a miniature version of the radio core of an active galaxy, these sources could move relativistically, changing their positions and brightnesses in weeks. If they represent structures in the 700-km-s^{-1} wind of ionized gas originating in the group of hot stars projected



The inner parsec of our Galaxy at 5 GHz radio frequency (Very Large Array map by M. Morris and F. Yusef-Zadeh). Most of the emission is from streams of ionized gas. The unusual point radio source Sgr A* is indicated. Also shown is the location of the cluster of hot young stars (IRS 16) from which emanates a high-speed wind which appears to be pushing material back from the star IRS 7 (E. Scrabyn, J. H. Lacy and J. M. Achtermann, personal communication), and perhaps holding back the dense ring of neutral molecular gas whose inner boundary is sketched.

duced by positron-electron annihilation was seen by low-resolution detectors, and was formerly ascribed to the Galactic Centre. But it has since been shown to originate many degrees away, probably by annihilation of positrons emitted by the radioactive ejecta of novae and supernovae, which may also excite the intense 6.7-keV iron line¹⁰. A possibly variable annihilation-line source¹¹ has not been definitively located, but if present is likely to be not Sgr A*, but an unusually bright source of high energy X- and γ -rays some 100 pc from the Galactic Centre¹².

As the other circumstantial evidence for black-hole powered activity in our Galactic Centre is explained away, the mystery of Sgr A* deepens. It resembles a very faint version of the radio cores of active galaxies. At the time of its discovery almost two decades ago¹³, this similarity suggested that it too represented a jet or disk about a supermassive accreting black hole. The absence of coincident infrared emission now makes this seem quite unlikely. Yet it is apparently stationary⁹, which implies that it is much heavier than a neutron star.

The discovery by Yusef-Zadeh, Morris

only 0.2 light year from Sgr A*, their motion could be detected in a decade. If they are not part of the wind, and simply orbit in the local gravitational potential, their motion will not be detectable for a century unless Sgr A* is the site of a supermassive black hole. If it is, unfortunately, the expected velocities are comparable to

- Kunieda, H., Turner, T., Awakai, H., Koyama, K., Mushotzky, R. & Tsusaka, Y. *Nature* **345**, 786–788 (1990).
- Kormendy, J. *Astrophys. J.* **325**, 128–141 (1988).
- Mould, J. et al. *Astrophys. J.* **339**, L21–L24 (1989).
- Richstone, D., Bower, G. & Dressler, A. *Astrophys. J.* **353**, 118–122 (1990).
- Yusef-Zadeh, F., Morris, M. & Ekers, R.D. *Nature* **348**, 45–47 (1990).
- Genzel, R. & Townes, C.H. *A. Rev. Astr. Astrophys.* **25**, 377–423 (1987).
- Sellgren, K. et al. *Astrophys. J.* **359**, 112–120 (1990).
- Allen, D.A. et al. in *The Center of the Galaxy*, Proc. IAU Symp. 136 (ed. Morris, M.) 513–515 (Kluwer, Dordrecht, 1989).
- Backer, D.C. & Sramek, R. in *The Galactic Center*, AIP Conf. Proc. 155 (ed. Backer, C.D.) 163–165 (AIP, New York, 1987).
- Koyama, K. et al. *Nature* **339**, 603–605 (1989).
- Leventhal, M. et al. *Nature* **339**, 36–38 (1989).
- Cook, W.R. in *The Center of the Galaxy*, Proc. IAU Symp. 136 (ed. Morris, M.) 581–585 (Kluwer, Dordrecht, 1989).
- Balick, B. & Brown, R.L. *Astrophys. J.* **194**, 265–270 (1974).
- Rees, M.J. *Nature* **333**, 523–528 (1988).
- Verbunt, F. in *Neutron Stars and their Birth Events* (ed. Kundt, W.) 179–218 (Kluwer, Dordrecht, 1990).
- Piotto, G., King, I.R. & Djorgovski, S. *Astr. J.* **96**, 1918–1924 (1988).

that of the 700-km-s⁻¹ wind, so the interpretation of such motion will not be unambiguous. The organization of the motion might allow the two possibilities to be distinguished, however.

If there is a supermassive black hole lurking quietly, it may someday tear apart a star which flies too close¹⁴. Our descendants could then see the centre of our Galaxy increase in luminosity by a factor of more than 10³. Even if there is no supermassive black hole, stars are still certainly being shredded — by each other. The density of stars in the inner parsec is so high that most stars there will have interacted with another at some time during their lives. The interactions range from passages in which tides heat or remove the outer layers of a star, to head-on collisions, resulting in total disruption or a merger of the two stars.

In some globular clusters, stars spend brief periods in cores with densities comparable to that of the inner parsec of the Galaxy. These clusters are distinguished by unusual numbers of X-ray sources and recycled pulsars, both believed to have been formed in collisions between neutron stars and ordinary stars¹⁵, and by unusual stellar populations¹⁶. How have

such things escaped detection in the Galactic Centre? Obscuration by dust makes determination of the properties of stars difficult, and dispersion by ionized gas can smear the pulses of short-period recycled pulsars beyond recognition.

The absence of hard X-ray sources, however, is real. This may indicate that neutron stars are largely absent from the Galactic Centre, as would be expected if most of the stars there were much heavier than the 1.4 M_⊙ of a neutron star (by contrast, in a globular cluster, the neutron stars are the heaviest bodies, and so sink to the dense centre). Collisions of older stars can create short-lived massive stars with unusual properties. These may evolve into black holes which can in turn disrupt stars and accrete the remains. Whether adorned by a 10⁶-M_⊙ black hole or not, the central parsec of our Galaxy, like the centre of a large city, is a crowded and dangerous environment. It will repay continued surveillance with lessons on the digestion and dismemberment of stars. □

E. S. Phinney is in the Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, California 91125, USA.

XERODERMA PIGMENTOSUM

A gene for tumour prevention

Richard D. Wood and Tomas Lindahl

WHEN Moritz Kaposi first described the disease xeroderma pigmentosum (XP), he noted the inherited nature of the syndrome and the skin tumours on sunlight-exposed portions of the body in young patients, but he could only speculate on the underlying cause of the disorder¹. Now, over one hundred years later, the complementary DNA sequence of a gene that is altered in one of the main forms of the disease is reported on page 73 of this issue².

Cells from this cancer-prone syndrome are defective in DNA excision repair³, exhibiting decreased removal of pyrimidine dimers and a variety of chemical adducts from damaged DNA. Even though XP is a rare, autosomally recessive disorder (usually occurring with a frequency of about 1 in 250,000), it displays substantial genetic heterogeneity, as demonstrated by cell-fusion experiments which currently define seven well-characterized complementation groups and an apparent 'variant' form⁴. Cells from the XP complementation groups have a defect in the incision step which initiates the excision-repair pathway^{5,6}. Complementation group A of XP accounts for one of the most prevalent and serious forms of the disease, representing approximately one-quarter of the patients tested.

Because XP cells are hypersensitive to

ultraviolet light, it was thought that isolation of the relevant genes by transfection of XP cells to radiation resistance would be a straightforward task, and a number of research groups took this approach. The experiments turned out to be more complicated than expected because spontaneous revertants appear in the recipient cell cultures, which can be easily mistaken for transfected cells, and also because human cells are more difficult to transfect stably with high-molecular-weight DNA than are rodent cells⁷.

Thus complementation by Tanaka *et al.*⁸ of the defective gene in cells of group A patients by transfection with mouse genomic DNA was a technical feat. The same group now describes the isolation and sequence of the human cDNA². The group A XP-correcting gene encodes a hydrophilic protein of relative molecular mass 31,000 (*M*_r 31K) which contains many charged residues. Interestingly, the protein displays a distinct zinc-finger motif, indicating that it interacts directly with DNA, presumably as part of an enzyme complex that makes an incision near damaged sites.

There is good evidence that the gene isolated by Tanaka *et al.* is the site of the primary defect in group A XP patients. The corresponding messenger RNA is greatly reduced in cell lines from several

severely affected Japanese patients, due to a G→C transversion at a splice-acceptor site. Certain group A patients with milder symptoms of the disease have normal levels of mRNA but carry a missense mutation in the cDNA sequence. Moreover, *in situ* hybridization locates the gene to human chromosome 9q34.1, an assignment that fully agrees with chromosomal transfer experiments performed recently by three separate research groups.

Announcement of the characterization of the group A XP-correcting gene appears almost simultaneously with a report by Weeda *et al.*⁴ that a previously cloned human DNA repair gene, *ERCC-3*, corresponds to the defective gene in the rare group B of XP⁴. The *ERCC-3* gene was initially isolated by its ability to correct an excision repair-defective Chinese hamster cell line⁹; the gene is located on human chromosome 2. Microinjection of the cDNA for *ERCC-3* into the nuclei of group B XP cells restored the ability of these cells to perform DNA repair synthesis after ultraviolet irradiation.

The cDNA sequence of the gene reveals an open reading frame for a polypeptide of *M*_r 89K, which again has features of a protein that interacts with DNA. The sequence contains characteristic amino-acid motifs that have been noted previously in a family of proteins with DNA helicase activity. Weeda *et al.* speculate that the primary function of the gene product may be to act as a DNA helicase, for example to unwind DNA locally at the site of a lesion, but it is also possible that the putative helicase activity may be a reflection of another function which involves DNA-dependent ATP hydrolysis. The inherited mutation identified in the defective group B XP gene investigated is also a transversion occurring at a G-C pair within a splice-acceptor site.

Several unusual cases of XP have been described in which the patients also exhibit the distinct symptoms of Cockayne's syndrome. The group B XP defect described by Weeda *et al.*⁴ belongs in this group. In its typical form, Cockayne's syndrome is also associated with sensitivity to ultraviolet light, but, by standard

1. Kaposi, M. *Wiener mediz. Jahrbücher* Oct., 619–632 (1882).
2. Tanaka, K. *et al. Nature* **348**, 73–76 (1990).
3. Cleaver, J. E. *Nature* **218**, 652–656 (1968).
4. Weeda, G. *et al. Cell* **62**, 777–791 (1990).
5. Cleaver, J. E. & Kraemer, K. H. in *The Metabolic Basis of Inherited Disease*, 6th edn. (eds Scriver, C. R. *et al.*) 2949–2971 (McGraw-Hill, New York, 1989).
6. Hansson, J. *et al. Nucleic Acids Res.* **18**, 35–40 (1990).
7. Mayne, L. V. *et al. Gene* **66**, 65–76 (1988).
8. Tanaka, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 5512–5516 (1989).
9. Weeda, G. *et al. Molec. cell. Biol.* **10**, 2570–2581 (1990).
10. Venema, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 4707–4711 (1990).
11. Arrand, J. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 6997–7001 (1989).
12. Chu, G. & Chang, E. *Science* **242**, 564–567 (1988).
13. Wood, R. D. *et al. Cell* **53**, 97–106 (1988).

methods of detection, cells are not appreciably defective in their ability to repair DNA. The cells do, however, have reduced excision-repair capacity for actively transcribed genes¹⁰. It seems likely that there is a partial biochemical overlap between these two clinical conditions, with spontaneously occurring mutations resulting in either XP or Cockayne's syndrome, or occasionally in both. By comparison, different mutations in human haemoglobin may cause several distinct diseases.

Attempts to clone other XP genes are now in progress¹¹, and an apparent loss of a separate protein that binds to damaged

DNA has been reported for the unusual XP complementation group E (ref. 12). In addition, a soluble cell-free system has been developed for functional assessments of purified XP gene products by biochemical complementation *in vitro*¹³. With some guidance from the extensive studies on *Escherichia coli* Uvr proteins and yeast *rad* mutants, it seems likely that we shall soon have a detailed understanding of the human defence mechanism against the powerful carcinogen sunlight. □

Richard D. Wood and Tomas Lindahl are in the Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, EN6 3LD, UK.

MITOSIS

The kinetochore in captivity

T. J. Mitchison

DURING mitosis, chromosomes interact with the spindle through specialized structures called kinetochores, to which spindle microtubules attach. The importance of kinetochores in chromosome segregation has been clear since their discovery by early cytologists¹, and recent studies of the kinetochore-microtubule interface have shown it to be the main site of force generation for chromosome movement². The major trend in mitosis research is now to identify the molecular components required for chromosome movement. But molecular biology alone could not easily address two old and very basic questions about the kinetochore: how do microtubules become attached to it, and what are the structural requirements for force generation? The first of these questions has now been answered by some experiments in the classic mould reported by Hayden, Bowser and Rieder in *Journal of Cell Biology*³.

Microtubule attachment has been the subject of vigorous, and sometimes acrimonious, debate for many years. After breakdown of the nuclear envelope in prophase of mitosis, kinetochores are initially devoid of attached microtubules. Within a few minutes microtubules are found connecting the centrosome and kinetochores, which may start many micrometres apart. Two hypotheses have been advanced to explain how this apparently difficult feat of connecting two small objects at a distance is accomplished. Either the microtubules are nucleated at the centrosome, and captured by kinetochores, or else they are nucleated at kinetochores, and only later interact with the spindle pole. The high density of microtubules in the forming spindle, and the complexity of their arrangements, made this issue difficult to decide by either light or electron microscopy.

The second question, that of the structural requirement for force generation, is

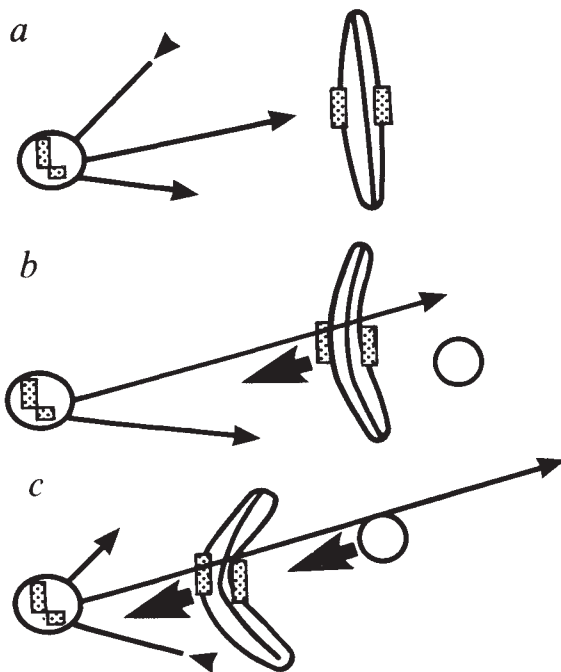
also a long-standing problem in mitosis, and its importance has been highlighted by recent developments in understanding the molecular mechanisms for intracellular transport. Motor proteins — kinesin, dynein and their kin — play a central role in transporting various structures over the microtubule lattice, and dynein is now known to be a component of kinetochores^{4,5}. Dynein may well turn out to be the motor for anaphase chromosome movement, but it remains possible that chromosomes are pulled polewards by microtubule depolymerization⁶. Struc-

tural evidence is crucial for distinguishing models of force generation: the depolymerization mechanism requires that the kinetochore must hold on to the end of the microtubule, whereas the motor-based mechanism does not.

These basic structural questions have been addressed in an elegant series of experiments in the laboratory of Conly Rieder. The experiments are very much in the tradition of mitosis experimentation, applying painstaking microscopic observation to the mitotic spindles of dividing cells in cultured newt-lung explants. Newts have the largest genome in the animal kingdom. They have correspondingly large cells, the flat morphology and large spindle size of which makes them ideal for light-microscopic observation of mitosis. Rieder and co-workers have used this culture in conjunction with modern video microscopy to draw some crucial conclusions about how the kinetochore works. Perhaps most important was their demonstration that kinetochores could *slide* polewards along microtubules early on in mitosis⁷. This work was discussed in an earlier News and Views article in conjunction with the discovery of dynein at kinetochores⁸. Here I will only reiterate how influential this study was, and how much was learned from the skilful and thoughtful application of approaches which many thought had already been mined for all they could deliver.

Hayden *et al.*³ have now pushed the optical microscopy a little further, using video contrast enhancement to bring out tiny details in the flat, clear newt spindles. They have succeeded in imaging individual microtubules as they grow out from mitotic poles and first come into contact with kinetochores. Contact is established as the microtubule grows out towards the periphery of the spindle, brushing past the side of the kinetochore. The kinetochore latches onto the microtubule lattice and the chromosome is jerked polewards. In some sequences a vesicle latches onto the same microtubule (see figure), and then follows the same microtubule track in towards the pole. Kinetochore and vesicle move at similar rates. Given the presence of dynein in kinetochores^{4,5}, the polewards movement of both kinetochores and vesicles is presumably driven by this motor protein.

The new observations lay to rest the controversy of how microtubules become attached to kinetochores, at least as far as typical animal spindles are concerned. They are nucleated



The process of microtubule attachment to the kinetochore. *a*, Several microtubules grow out from, and shrink back towards, the spindle pole during prometaphase. *b*, One microtubule makes a lateral contact with a kinetochore. The kinetochore attaches to it and starts moving polewards. *c*, A vesicle makes a lateral contact with the same microtubule and also moves polewards. (Adapted from the video micrographs of Fig. 4 in ref. 3.)

at the centrosome, and captured by kinetochores. It is not yet clear whether all capture events involve lateral interactions initially, or whether direct end-on capture is also possible — this may depend simply on the geometry of the initial interaction. It is with obvious relish that the authors report that they have confirmed the "century old hypothesis of van Beneden and Neyt that 'kinetochore fibres are simply those astral rays that come into connection with chromosomes'".

I allied myself with the kinetochore capture school, and thus against the kinetochore nucleation school, in 1984 (ref. 9). Kinetochore capture provided a biological role for the then bizarre dynamic instability phenomenon exhibited by microtubules *in vitro*. Kirschner and I hypothesized that, during mitotic spindle assembly, microtubules grow out from centrosomes and shrink back if they do not come into contact with kinetochores. Dynamic instability causes continuous growth and shrinkage of microtubules, allowing them to probe through the cytoplasm and thus increase the probability of finding kinetochores. This hypothesis has now been confirmed.

Like all good observations, those of Hayden *et al.*³ pose as many questions as they solve. Chromosomes move much more slowly once their kinetochores establish mature, end-on connections to microtubules, and the structural and molecular basis of this velocity regulation is unclear. Also unclear is what happens to the minus end of the microtubule at the mitotic pole following kinetochore capture of the plus end. Photoactivation marking experiments have revealed that, in mature kinetochore fibres, microtubules depolymerize slowly at their polewards ends, while polymerizing at the kinetochore¹⁰. Does this imply release of the microtubule minus end from the nucleation site, and if so why? There are many mysteries remaining as to the mechanism of spindle formation and chromosome movement, and it is likely that careful observations and structural analyses, such as those described here, will continue to make an important contribution to advances in the field. □

T.J. Mitchison is in the Department of Pharmacology, University of California, San Francisco, California 94143, USA.

Open and shut case for anions

Philip C. H. Mitchell

CHEMISTS have long been fascinated by molecular structures with a central cavity which can encapsulate a molecule or ion. Such cavities are found in solids such as zeolites (Fig. 1). Molecules such as crown

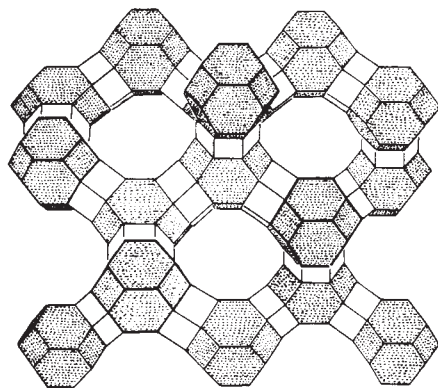


FIG. 1 A typical zeolite structure (faujasite) showing cavities which can accommodate guest molecules. Al and Si atoms are at the apices of the polyhedra and are linked by oxygen atoms (shown as lines). (From ref. 7.)

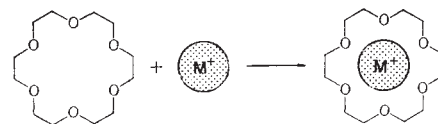


FIG. 2 A crown ether (18-crown-6) encapsulating a metal ion. The O-atoms are linked by $-\text{CH}_2-\text{CH}_2-$ units. (From ref. 8.)

ethers (Fig. 2), and analogous natural products like the antibiotic valinomycin, are organic compounds with a particular ability to encapsulate electropositive cations (Na^+ , K^+ and so on) and render them soluble in organic solvents. Until very recently, purely inorganic analogues of the crown ethers had yet to be prepared. But now Achim Müller and colleagues in Germany report¹ a series of water-soluble vanadium oxides with encapsulated ions (chloride, bromide, iodide, carbonate). The compounds are easily made by heating vanadium pentoxide or a vanadate with a water solution of a hydrazinium (N_2H_5^+) salt and an anion. The work follows an earlier report from V.W. Day and colleagues² of the synthesis of a vanadium oxide inclusion complex, $[(\text{CH}_3\text{CN})\text{V}_2\text{O}_7]^{4-}$, which is soluble in acetonitrile. The new compounds are especially interesting because the encapsulated species are anions, whereas the guest species in most inclusion compounds are neutral molecules or cations.

In this branch of chemistry, a key idea is the structural match, in size and shape, of host and guest — leading to the property of selectivity. A zeolite with an appropriate channel and cavity size will admit a straight-chain hydrocarbon but exclude

a bulkier branched-chain compound. Similarly, a particular crown ether, like the valinomycin antibiotic, may encapsulate and transport K^+ in preference to Na^+ because the cavity best accommodates the K^+ ion. The soluble vanadium oxides have a cavity size and charge which enables them to encapsulate halide and other anions.

Müller and colleagues' soluble vanadium oxides are hollow spheres built from linked VO_5 units (Fig. 3). The spheres are hollow because of the propensity of vanadium (v), and more especially (iv) (formed by hydrazine reduction of vanadate), for fivefold coordination. In these VO_5 species, vanadium (iv) is at the centre of a square pyramid of oxygen atoms. With this structure, one of the six potential coordination sites (with octahedral symmetry) of vanadium is empty and remains so in the soluble oxide. The encapsulated anions are located symmetrically in the spheres' hollow centres.

The structure of the carbonate inclusion compound $(\text{Li}_7[\text{V}_{15}\text{O}_{36}(\text{CO}_3)]\cdot\text{aq})$ is especially fascinating (Fig. 3). There are two equivalent positions within the hollow sphere where the triangular carbonate may occupy, the three carbonate oxygens interacting with one or other of two sets of three framework vanadium atoms. The vanadium oxide framework relaxes towards the carbonate ion, so that the final minimum-energy structure is a compromise between energy loss attributable to distortion of the oxide and energy gain due to the attraction of vanadium for carbonate. The six interacting vanadium atoms are displaced towards the centre of

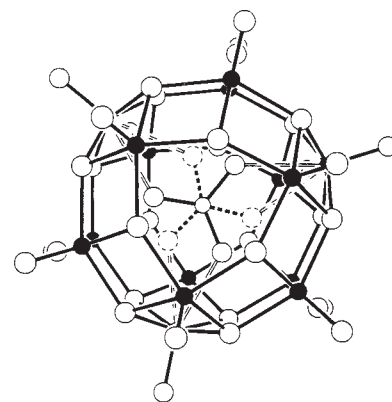


FIG. 3 Structure of Müller and colleagues' soluble vanadium-oxide/carbonate inclusion complex (from ref. 1). Oxygen and carbon (of the central carbonate) are shown as open circles and vanadium as closed circles. Each sphere encapsulates only one carbonate; the broken lines represent the alternative, equivalent carbonate position. Oxygens of the carbonate are shown linked to framework vanadium atoms.

1. Wilson, E.B. *The Cell in Development and Inheritance*, 2nd edn (Macmillan, New York, 1911).
2. Mitchison, T.J. *A. Rev. Cell Biol.* **4**, 527–549 (1988).
3. Hayden, J.H., Bowser, S.S. & Rieder, C.L. *J. Cell Biol.* **111**, 1039–1045 (1990).
4. Pfarr, C.M. *et al. Nature* **345**, 263–265 (1990).
5. Steuer, E.R. *et al. Nature* **345**, 266–268 (1990).
6. Koshland, D.E. *et al. Nature* **331**, 499–504 (1988).
7. Rieder, C.L. & Alexander, S.P. *J. Cell Biol.* **110**, 81–96 (1990).
8. Vallee, R. *Nature* **345**, 206–207 (1990).
9. Mitchison, T.J. & Kirschner, M.W. in *The Molecular Biology of the Cytoskeleton* (eds Borisy, G.G. *et al.*) 27–44 (Cold Spring Harbor Laboratory, New York, 1984).
10. Mitchison, T.J. *J. Cell Biol.* **109**, 637–652 (1989).

the sphere — towards the carbonate — by about 19 picometres. The V–O (carbonate) distance is then about 230 pm, rather longer than a V–O single bond (200 pm). The guest–host interaction (which I prefer to describe as ion–dipole rather than covalent) thus causes the host to adjust to the guest. The structure is significant in the light of current computational studies of zeolites³ (including our own work⁴) which indicate that a zeolite framework, for example ZSM-5, should relax to accommodate a guest molecule, such as

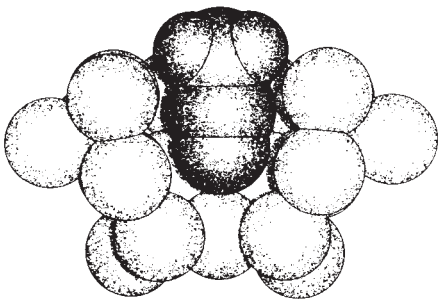


FIG. 4 A basket-shaped soluble vanadium oxide inclusion complex $[(CH_3CN) V_{12}O_{32}]^{4-}$ (from ref. 2). In this space-filling representation framework oxygens are shown as spheres surrounding the CH_3CN guest molecule. Note how the bulky CH_3 group (top) of the acetonitrile guest, which functions as the template molecule, projects from the basket preventing the completion of a vanadium oxide sphere.

methanol⁵. Müller and co-workers' new compounds, well characterized molecular analogues of zeolites, will probably provide a field day for solid-state computational chemists.

Without the captured anions, the spherical structures of the new oxide complexes would presumably not develop, so that evidently we have here examples of 'template' reactions: the guest species provides a template for the structure of the host. Such reactions are well known in coordination chemistry, in which a metal ion serves as a template for the organic synthesis: an example is copper phthalocyanine⁶. Analogously, in solid-state oxide chemistry, a template reaction is used to synthesize zeolitic aluminium phosphates⁶. Alumina, phosphoric acid and an amine (the template) are heated in water under pressure. The aluminium phosphate is built around the organic amine, the cavity size being determined by the size of the amine. To remove the organic amine and generate an empty cavity the aluminium phosphate amine must be heated to 600 °C. What we do not yet know is whether it is possible to remove the captured anions from Müller and colleagues' oxides, creating an empty cavity while retaining the molecular structure.

Zeolites have important applications as molecular sieves and heterogeneous catalysts. Confined within the cavity, a molecule may encounter other molecules or be activated by interaction with the

walls (a C–H bond of methanol, for example is stretched in ZSM-5 (ref. 3)). The cavity provides a reaction space rather as the active site of an enzyme does. Could the soluble vanadium oxides have analogous applications in homogeneous solution chemistry? Vanadium oxides are known to be selective oxidation catalysts, and the new soluble vanadium oxides, like zeolites, provide a reaction space. With vanadium one also has the possibility of redox chemistry involving the vanadium framework. What is crucial is that the reactant molecules or products should be able to pass in and out of the cavity: perhaps a porter is needed as well as a host.

From this point of view, Day's soluble vanadium oxide inclusion complexes² are especially attractive. The vanadium oxide structure builds around the acetonitrile template but is prevented from closing by

CELL RESPIRATION

Catalytic intermediates

Mårten Wikström and Gerald T. Babcock

ALL higher forms of life depend on cell respiration to release the energy of food-stuffs by their oxidation with O_2 . On page 89 of this issue¹, Han, Ching and Rousseau make an important contribution to our understanding of this process by reporting the physical verification of the structures of two of the catalytic intermediates involved.

The reduction of O_2 in cell respiration is common to eukaryotes and aerobic prokaryotes, and is catalysed by a family of haem–copper proteins, the cytochrome oxidases. In eukaryotes there is only one such enzyme, cytochrome *c* oxidase,

the bulky methyl group. The vanadium oxide then forms a hemisphere or basket (Fig. 4) in and out of which one can easily imagine molecules passing. □

Philip C. H. Mitchell is in the Department of Chemistry, University of Reading, Whiteknights, PO Box 224, Reading RG6 2AD, UK.

1. Müller, A., Penk, M., Rohlfing, R., Krickemeyer, E. & Döring, J. *Angew. Chem. Int. Edn Eng.* **29**, 926–927 (1990).
2. Day, V.W., Klemperer, W.G. & Yaghi, O.M. *J. Am. chem. Soc.* **111**, 5959–5961 (1989).
3. Vetrivel, R., Catlow, C.R.A. & Colbourn, E.A. *J. phys. Chem.* **93**, 4594–4598 (1989).
4. Drew, N.G.B., Jutson, N.J., Mitchell, P.C.H. & Wass, S.A. *Polyhedron* **8**, 1817–1818 (1989).
5. Cotton, C.A. & Wilkinson, G. *Advanced Inorganic Chemistry* 5th edn 344–347; 353–354 (Wiley, New York, 1988).
6. Wilson, S.T., Lok, B.M., Messina, C.A., Cannan, T.R. & Flanigen, E.M. *J. Am. chem. Soc.* **104**, 1146–1147 (1982).
7. Wells, A.F. *Structural Inorganic Chemistry* 4th edn, 830 (Oxford, 1975).
8. Hiraoka, M. *Crown Compounds*, 68 (Elsevier, Amsterdam, 1982).

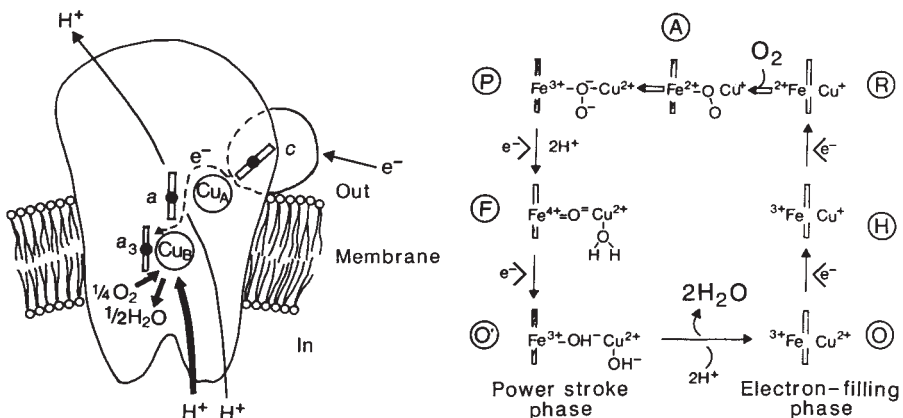


FIG. 1 (left), Schematic diagram of cytochrome *c* oxidase (cytochrome aa_3) in the inner mitochondrial or bacterial cell membrane. Electrons are transferred from the haem of cytochrome *c* by way of Cu_A and haem *a*, to the bimetallic O_2 -reduction centre consisting of haem a_3 and Cu_B . The reaction is linked to uptake of $1 H^+/e^-$ to form water from reduced oxygen, and to pumping of $1 H^+/e^-$ across the membrane. FIG. 2 (right) Proposed mechanism of O_2 reduction in cell respiration (modified from refs 9 and 10). Seven intermediate states of the bimetallic haem a_3 – Cu_B reaction centre are shown (O, oxidized; H, half-reduced; R, reduced; A, oxy; P, peroxy; F, ferryl; O', ferrichydroxide). Four electron transfers to the centre from cytochrome *c* (by way of Cu_A and haem *a*, Fig. 1) are indicated by $\leftarrow e^-$. Proton pumping (not shown here) is coupled only to the electron transfers of the 'power stroke' phase¹¹.

iates in the oxidase reaction.

The reduction of O_2 to water is not a trivial chemical process, but may involve superoxide (O_2^-), peroxide (H_2O_2) and hydroxyl radical ($HO\cdot$) as intermediates, all of which are potentially harmful. Indeed, much research is being devoted to understanding the biochemistry of these 'oxygen radicals' and the damage they cause to cells. Because cell respiration accounts for 90 per cent or more of the consumption of O_2 in the biosphere, it is crucial that these compounds are not released to any measurable extent in the process. Moreover, owing to its unique physicochemical properties, atmospheric O_2 is relatively unreactive at ambient temperatures — which is fortunate, as otherwise organic material would burn spontaneously and life on Earth would be jeopardized. But how does cytochrome oxidase activate O_2 so that it is quickly reduced in a well-controlled process that avoids the release of toxic intermediates?

Central to setting the stage for description of the catalytic cycle was the discovery² in 1969 of the bimetallic haem iron-copper centre and the subsequent confirmation of its identity by magnetochemical techniques³⁻⁵. In 1975, the extension to low temperatures⁶ of the pioneering flow-flash kinetic techniques developed earlier⁷ led to the identification of the first catalytic intermediate of the oxidase reaction — the ferrous-oxy form analogous to oxyhaemoglobin (Fig. 2, A). Subsequently, this intermediate was identified at room temperature⁸.

In 1981, two discrete intermediates that seemed to have peroxy and ferryl structures (Fig. 2, P and F) were identified by optical spectroscopy⁹ during partial rever-

sal of the catalytic mechanism. Later, evidence was obtained¹⁰ for the ferryl intermediate in the forward, reaction, providing important information on its kinetics of formation and chemical reactivity.

A proposed^{9,11} model for the catalytic cycle is shown in figure 2. Coupling to proton translocation occurs only in the 'power stroke' transfer of electrons to the peroxy and ferryl intermediates. Raman spectroscopy, which can provide positive identification of these proposed intermediates, was first applied in 1984 to the cytochrome oxidase/ O_2 reaction¹². Recently, the existence of the primary oxy intermediate has been verified by this technique¹³⁻¹⁵, and the ferryl species¹ has been identified independently in two

other laboratories^{16,17}. The catalytic mechanism shown in the Fig. 2 is now supported by firm physical data.

One remaining problem is the failure, so far, to demonstrate the existence of a ferric peroxy species (Fig. 2, P) the forward reaction of the fully reduced enzyme. It is possible that the formation of this species is rate-limiting. If so, fast electron transfer would lead to its rapid decomposition to the ferryl state, possibly by way of some ferrous-peroxy-cupric intermediate¹⁸. □

Mårten Wikström is in the Department of Medical Chemistry, University of Helsinki, Siltavuorenpenger 10A, SF-00170 Helsinki, Finland. Gerald T. Babcock is in the Department of Chemistry, Michigan State University, East Lansing, Michigan 48824-1322, USA.

MANTLE COMPOSITION

The endmember stew

Richard W. Carlson

ISOTOPIC shifts in elements with long-lived radioactive parents document chemical variations in the Earth's mantle that have been preserved for periods of at least 10^8 – 10^9 years — roughly the timescale of convective overturn in the mantle. Relying primarily on data from oceanic volcanic rocks, because of their freedom from contamination by the old materials of the continents, correlated variations in strontium, neodymium and lead isotope composition have been interpreted as mixing arrays between 4 or 5 isotopically distinct components in the mantle^{1,2}. Isotope data for most oceanic islands rarely reach these 'endmember' components, however, raising the issue of whether the endmembers represent volumetrically significant reservoirs or instead reflect only hypothetical extremes of chemical fractionation or age of preservation. Data reported on page 59 of this issue³ by Barling and Goldstein allow this question to be addressed directly for the first time.

The most voluminous of all oceanic volcanism, that occurring along the worldwide system of oceanic ridges, defines the predominant endmember of the oceanic isotope data, the MORB (mid-ocean-ridge basalt) source. Ocean-ridge basalts are depleted in the so-called incompatible elements that preferentially partition into the magma during partial melting of the mantle. The radiogenic Nd and non-radiogenic Sr isotope compositions that typify the MORB endmember show that the 'depleted' character of MORB is not a recent phenomenon, but instead results from melt removal from the mantle that occurred at least 10^8 years ago. Most probably the incompatible-element depletion of the MORB mantle was caused by the extraction of continental and oceanic crust.

In spite of the ubiquitous occurrence of depleted MORB at oceanic spreading ridges, ocean-island volcanism always appears to involve greater enrichment of incompatible elements. Isotope compositions for island volcanism also trend away from the MORB endmember towards a variety of points whose isotope compositions are indicative of long-term (10^8 – 10^9 year) incompatible-element enrichment, at least as compared to MORB. Over the past decade, most of these enriched components have been put forward as representing various near-surface, incompatible-element enriched materials such as oceanic crust or continental crust that have been recycled and imperfectly mixed back into the mantle by plate subduction and mantle convection⁴.

With the acceptance of recycling as the main way of creating chemical heterogeneity in the mantle, the discussion has shifted to the nature of the storage of enriched components in the mantle and their reintroduction into the sources of ocean-island volcanism. At one extreme, enriched components are proposed to represent 'plums' of material scattered in a mantle-wide matrix of MORB-type host. Alternatively, the enriched components may reside in discrete storage zones, for example, the D''-layer at the base of the mantle, to be reinjected into the upper mantle as plumes rising through and mixing with a MORB-dominated upper mantle.

The way that Barling and Goldstein's data³ impinge on this argument is that, serendipitously, the lavas from Heard Island in the southern Indian Ocean define a strongly curved isotope array that allows clear resolution of the isotope characteristics of the mixing endmembers. In the linear trends shown by most

- Han, S., Ching, Y.-c. & Rousseau, D. L. *Nature* **348**, 89–90 (1990).
- van Gelder, B. F. & Beinert, H. *Biochim. biophys. Acta* **187**, 1–24 (1969).
- Falk, K.-E., Vänngård, T. & Ångström, J. *FEBS Lett.* **75**, 23–27 (1977).
- Tweedle, M. F., Wilson, L. J., Garcia-Iniguez, L., Babcock, G. T. & Palmer, G. J. *biol. Chem.* **253**, 8065–8071 (1978).
- Moss, T. H., Shapiro, E., King, T. E., Beinert, H. & Hartzell, C. R. *J. biol. Chem.* **253**, 8072–8073 (1978).
- Chance, B., Saronio, C. & Leigh, J. S. Jr *J. biol. Chem.* **250**, 9226–9237 (1975).
- Gibson, Q. & Greenwood, C. *Biochem. J.* **86**, 541–554 (1963).
- Orii, Y. *J. biol. Chem.* **259**, 7187–7190 (1984).
- Wikström, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4051–4054 (1981).
- Witt, S. N. & Chan, S. I. *J. biol. Chem.* **262**, 1446–1448 (1987).
- Wikström, M. *Nature* **338**, 776–778 (1989).
- Babcock, G. T., Jean, J. M., Johnston, L. N., Palmer, G. & Woodruff, W. H. *J. Am. chem. Soc.* **106**, 8305–8306 (1984).
- Varotsis, C., Woodruff, W. H. & Babcock, G. T. *J. Am. chem. Soc.* **111**, 6439–6440; **112**, 1297 (1989).
- Han, S., Ching, Y.-c. & Rousseau, D. L. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2491–2495 (1990).
- Ogura, T., Takahashi, S., Shinzawa-Itoh, K., Yoshikawa, S. & Kitagawa, T. *J. Am. chem. Soc.* **112**, 5630–5631 (1990).
- Varotsis, C. & Babcock, G. T. *Biochemistry* **29**, 7359–7362 (1990).
- Ogura, T., Takahashi, S., Shinzawa-Itoh, K., Yoshikawa, S. & Kitagawa, T. *J. biol. Chem.* **265**, 14721–14723 (1990).
- Blair, D. F., Witt, S. N. & Chan, S. I. *J. Am. chem. Soc.* **107**, 7389–7399 (1985).

oceanic islands, it is not possible to discern whether the mixing endmembers lie at or beyond the extremes of the actual data. The linear data arrays usually are consistent with mixing between MORB and one of the several enriched components, even though the actual data may not reach either endmember. The hyperbolic arrays seen in the Heard data, however, do not pass through the MORB endmember, indicating that the depleted Heard component cannot be as depleted as MORB.

Most clearly, this rules out a model in which the Heard array is generated by the melting of a plum of enriched material embedded in a MORB matrix. Similarly, it also excludes the possibility of mixing between a plume of enriched material rising through a mantle composed primarily of MORB.

The significance of this result is that it shows that actual volumes of mantle exist which lie not at the extremes of the isotope range observed for oceanic volcanic rocks, but within it. This has been suspected for some time because of the predominance of lavas with isotope compositions falling within a narrow subset of the total oceanic array. The local concentration of data was even given the name PREMA (prevalent mantle) by Zindler and Hart² but it was never possible to resolve whether PREMA was a distinct component on its own or was itself simply the most likely mixture of MORB and some enriched component. Taken back one step, both PREMA and the depleted Heard component may indeed be created by mixing between MORB and a more enriched component.

The Heard data, however, demonstrate a hierarchy to the mixing — the intermediate mixture identified as the depleted Heard component must have existed as a discrete component in the mantle that was then mixed with additional enriched material during the genesis of the Heard volcanic system. The data suggest that the mantle is not composed of endmember components that mix only during the events leading up to ocean island volcanism. Instead, continuous mixing within the mantle, driven by convective motion, probably results in a range of compositions where, as Barling and Goldstein state, the extremes "reflect the most fractionated examples or the longest-surviving remnants of mantle fractionation processes". □

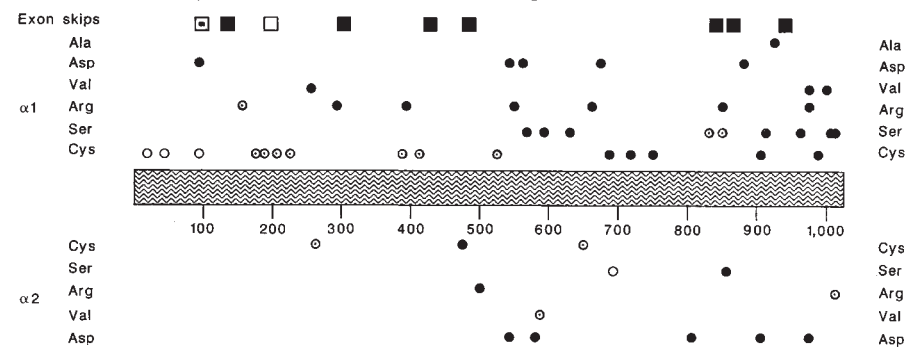
Richard W. Carlson is in the Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road N.W., Washington DC 20015, USA.

Bone disease cracks genetics

Bryan Sykes

SIDE-stepping the usual slog along the chromosome from linked marker to mutant gene has its price. When the inherited bone disease osteogenesis imperfecta was first linked to the favourite candidates, the type-1 collagen genes, all dramatic potential for scripting a long-running soap opera was lost. Gone were the opportunities for brief flirtations with different markers, each in turn discarded

because it contains two different subunits. Two out of the three are identical (the $\alpha 1$ chains) and the sequence of the third ($\alpha 2$) is very similar. They are encoded at the non-syntenic loci *COL1A1* and *COL1A2*. Folding of first the individual α chains themselves and then the triple superhelix places every third residue of all three chains at the axial centre of the molecule. So tight is the packing that only glycine,



Axial glycine (circles) and exon-skip (squares) mutations in the genes for the $\alpha 1$ (upper) and $\alpha 2$ (lower) subunits of type-1 collagen. The scale indicates helix residue number. Shading indicates the OI phenotype from lethal (solid) through severe (dot) to mild (open).

for a better-looking rival in an atmosphere of secrecy, treachery and betrayal. Lost were the chances of leading the audience step by panting step to the ultimate climax — the isolation of the gene itself. With osteogenesis it was more like marrying the girl next door. But now the *paparazzi* have moved on to higher things and chromosome-walking is beginning to look decidedly eighties, the advantages of an acquaintance built up over the years are beginning to show. Osteogenesis imperfecta (OI or brittle bone disease) is a dominant disease *par excellence* and is proving to be an excellent vehicle for unravelling the knotty problems of variable penetrance and expression associated with dominant traits. Among the most exciting issues to emerge at a recent meeting* was that somatic mosaicism might be a much more widespread and important factor in expression than had hitherto been appreciated.

Osteogenesis imperfecta is a disadvantageous trait, so the population frequency (about 5,000 cases in Britain and a similar proportion elsewhere) is maintained by recurrent mutations counterbalanced by selection. Why then is the mutation rate so high? The answer lies in the structure of the collagen gene-protein system itself. The heart of the collagen molecule is its long triple-helical domain of 1,014 amino acids, with the three subunit α -chains intimately entwined along the entire length. Each type-1 collagen mol-

ecule contains two different subunits. Two out of the three are identical (the $\alpha 1$ chains) and the sequence of the third ($\alpha 2$) is very similar. They are encoded at the non-syntenic loci *COL1A1* and *COL1A2*. Folding of first the individual α chains themselves and then the triple superhelix places every third residue of all three chains at the axial centre of the molecule. So tight is the packing that only glycine,

because it has no side chain, can occupy that position without major disruption. So, every third residue is glycine and each α chain has 338 of them. Substitution of axial glycines had always been seen as likely disruptive mutations but their frequency in OI has surprised most people. They were first noticed because cysteine, one of the eight possible substitutions brought about by a single base change in the glycine codon, introduces a new disulphide bond which can be detected by SDS electrophoresis. Mismatching techniques using either chemical cleavage (J. Bateman, University of Melbourne; M. Willing, University of Washington; M. Mottes, University of Verona), RNase (J. Marini, NIH, Bethesda) or, for the well-endowed, blanket sequencing (D. Prockop, Jefferson Medical College, Philadelphia), have now eliminated the need to detect a protein abnormality. By the end of the meeting the cumulative total of glycine substitutions had gone up to 53 with 40 in the $\alpha 1$ and 13 in the $\alpha 2$ chains (see figure). Cysteine is still found to be the most commonly substituted amino acid, probably because of a mixture of the bias in identifying residues, glycine codon usage and mutation preference. Six out of the eight amino acids that are within reach of a single base substitution of the GGN glycine codon have now been found. The exceptions are glutamic acid (GA(A/G)) and tryptophan (TGG). No doubt both will be identified in time although with only three GGG glycine codons in each

1. White, W. M. *Geology* **13**, 115–118 (1985).

2. Zindler, A. & Hart, S. R. A. *Rev. Earth planet. Sci.* **14**, 493–571 (1986).

3. Barling, J. & Goldstein, S. L. *Nature* **348**, 59–62 (1990).

4. Hofmann, A. W. & White, W. M. *Earth planet. Sci. Lett.* **57**, 421–436 (1982).

*4th International Conference on Osteogenesis Imperfecta, Pavia, Italy, 16–19 September 1990.

chain, Gly-Trp changes are bound to be rare.

There is a strong preference for mutation of the first rather than the second base of the glycine codon (41/53) and a marginal preference for mutations in CpG dinucleotides which are frequent targets in other genes. For a CpG pair to involve a glycine codon, the final base of the preceding codon must be a C and, although this is unusual (47/591 (8 per cent) in genomic DNA), 8 of the 41 first-base substitutions (19.5 per cent) occur in CpG pairs.

The enormous variation in the OI phenotype is one of its most intriguing features. It varies from forms that are lethal in the perinatal period, through crippling but non-lethal, to a comparatively mild disease where there is only a marginally increased tendency for bones to fracture. Examples from the entire range are found among the axial glycine mutants, so the question arises: how can the dramatically different effects of what are, on the face of it, fairly similar mutations be best explained? The three variables to play with are chain ($\alpha 1$ or $\alpha 2$), the substituted residue and its position.

Taking position first, even three years ago (see my previous News and Views¹ "Genetics cracks bone disease") it looked as though glycine substitutions had a more severe effect on the phenotype when they occurred towards the C-terminal end of the helix. By and large, that pattern has been adhered to by subsequent mutations found in the $\alpha 1$ chain, though there are exceptions. The correlation in $\alpha 2$ is much weaker. For instance a Gly-Cys mutation at position 646 was found in a fairly mild case, whereas an upstream Gly-Cys mutation at position 259 produced a much more severe phenotype (R. Wenstrup, Duke University) and at position 472 it was lethal. Generally, the $\alpha 1$ mutants are more damaging than their equivalents in the $\alpha 2$ chain. Nobody at the meeting disagreed with the reasoning that because there were two $\alpha 1$ chains and one $\alpha 2$ in each molecule, the simple stoichiometry of their construction would leave only one-quarter of molecules without a mutant chain if it were an $\alpha 1$ but one-half of molecules free of a mutant subunit if it were an $\alpha 2$.

The substituted residue also has some correlation with phenotype, so bulky aspartate is lethal wherever it occurs in the axial position, whereas arginine and cysteine are comparatively mild towards the N terminus. The pattern was all beginning to look a bit cosy until two Gly-Ser mutations at positions 832 and 844 in $\alpha 1$, which should have been lethal, were described from severely disabled but nonetheless surviving patients (P. Byers, University of Washington, Seattle). Rationalists blamed the presence of local 'permissive' regions for helix unfolding, but there doesn't seem to be anything

remarkable about the sequences surrounding positions 832 or 844.

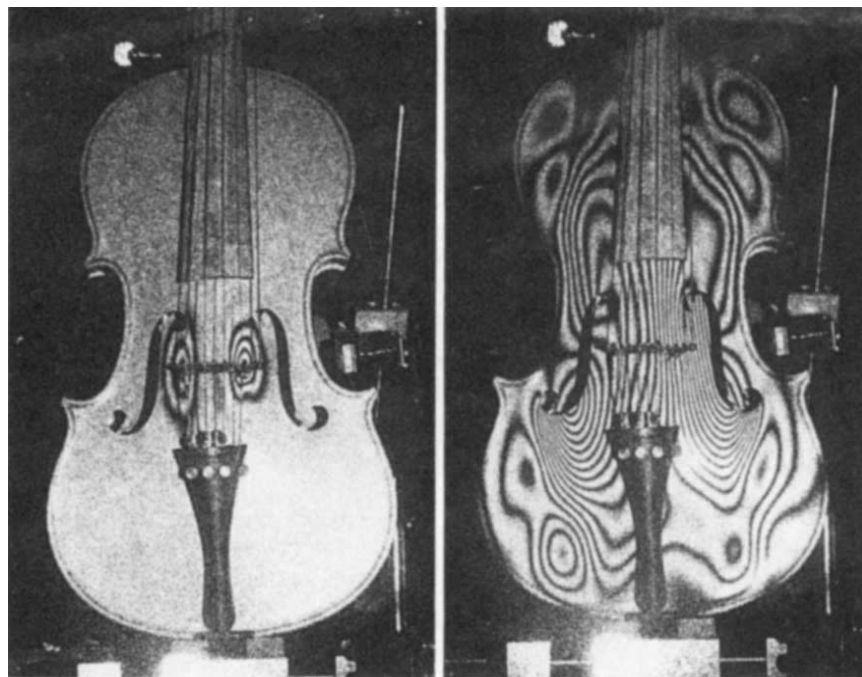
That some patients with the more N-terminal mutations lacked fractures, but still had overall osteoporosis and scoliosis, tempted some to blur the boundary between rare and common disease by claiming an OI-style mutation ($\alpha 2$ Gly-Arg 661) for a patient not with OI but with post-menopausal osteoporosis (D. Prockop). If similar mutations really are a frequent cause of such a common disease then the implications are obviously enormous. Until we know, there is a danger that such claims and others like it (for example Arg-Cys 591 in the related type-II cartilage collagen in osteoarthritis-chondrodysplasia²) might be mere semantic exercises.

Axial glycines apart, the other feature of collagen that makes it a sitting duck for mutation is the hundred or so intron-exon boundaries that have to be kept in good shape. The trickle of deletion mutants being discovered three years ago has become, if not a flood, then a steady stream. Multi-exon deletions seem to occur rarely, always result in a lethal phenotype and are true genomic deletions. More commonly, single exons are

missing from the messenger RNA, not as a result of genomic deletions but because of mutations in the splice-donor or acceptor sites. Those readers reared on globin might expect such errors to result in a failure to excise the mutant intron or the activation of cryptic sites. Not so for collagen. What happens is neat excision of the exon and its flanking introns, perhaps by a mechanism (D. Rowe, University of Connecticut) based on a recent model for exon definition³. After interaction at the 3' end, correct assembly of the spliceosome depends on a donor site being within 300 base pairs (bp) downstream. If the wild-type site on the other side of the exon is mutated and no other site is found within 300 bp then that interval is not recognized as an exon. Processing then excises the whole stretch between the intact upstream donor and downstream acceptor sites, including the now invisible exon.

Nine single-exon skips due to boundary mutations were described in $\alpha 1$ and seven in $\alpha 2$. Because the triplet frame is left intact, molecules can form and the $\alpha 1$ and $\alpha 2$ stoichiometry effect comes into play. All the $\alpha 1$ skip mutations are effectively lethal (a patient carrying one is still alive thanks only to intensive care), with the

Good vibrations from a violin



THE most characteristic part of a musical note, musicians hold, is the initial burst of sound: deprive a note of this initial waveform, and many instruments become indistinguishable. Nils-Erik Molin and colleagues (*J. acoust. Soc. Am.* 88, 2479-2481; 1990) have studied this moment for a violin by striking the instrument's 'bridge' (centre) with a light pendulum and, using laser interferometry, watching how the transient waveforms so stimulated propagate through its body. The figures above, taken 100 and 450 microseconds after impact, show the dipolar character of the waves generated at the two 'feet' of the bridge, and how the high frequencies quickly arrive at the 'f-holes' cut into the belly of the violin. The slight asymmetry in the dipolar waveform is attributable to subtle differences in the tension at each foot. □

exception of two, one of which gives rise to a very unstable message, and to a product that does not get into collagen molecules and cannot therefore exert a disruptive influence; the resulting phenotype is mild. In $\alpha 2$, only two lethal skip mutations have been found. The boundary changes are usually straightforward single-base substitutions, but sometimes the splice sites are removed as part of a longer deletion (A. Nicholls, Clinical Research Centre, London; A. Superti-Furga, University of Zurich).

As if to reassure traditionalists, two splicing mutants were described that did behave like globin. A mutation in the donor site of *COL1A1* intron 24 leads not to skipping of the upstream exon but to retention of the mutated intron in the processed RNA (D. Rowe). The rationale was that intron 24 is short enough for the donor site beyond exon 25 to be within reach, so that the whole stretch (exon 24, intron 24 and exon 25) is defined as a single exon and not spliced out. The RNA is unstable and is retained in the nucleus, so the phenotype is mild. For $\alpha 2$, an alternative donor site is close enough to be activated when the normal site is mutated, leading to an 18-bp insertion in the mRNA. This time the message is stable and cells synthesize an $\alpha 2$ chain with six non-helical residues between positions 585–586 (R. Wenstrup).

There were several nice touches outside the 'big two' mechanisms of glycine substitution and exon skipping. In one severe non-lethal case a 3-bp deletion removed the codon for $\alpha 1$ valine 255 (K. Molyneux, University of Leicester). Downstream folding of the helix should be out of phase in this mutant. Another deletion removed a single Gly-Ala-Pro triplet from $\alpha 1$ exon 43, presumably allowing in-phase helix folding to continue to the N terminus but nonetheless resulting in a lethal phenotype (G. Wallis, University of Washington, Seattle).

Although these examples show convincingly that there is enough potential within the structure of the two collagen genes for mutations to express the entire range of phenotype, the most interesting mechanism for introducing variation, and one with far-reaching implications for genetics in general, came from the study

of sibling recurrences in the lethal form of OI. Lethal OI was originally thought to be an autosomal recessive disease because of the occurrence of multiply affected siblings among the offspring of unaffected parents. Over the years a few pedigrees came to light in which there were affected half-sibs (that is, siblings with only one parent in common). The isolation of the causal mutation showed that they were due to the same sort of dominant mutations as the isolated cases and that the common parent was a germ-line mosaic. It now turns out that, for each of the eight cases of recurrences in sibs and half-sibs that have been studied, one parent is not only a germ-line but also a somatic mosaic (D. Cohn, Cedars-Sinai, Los Angeles) with the mutant allele present in up to 80 per cent of lymphocytes and 100 per cent of cultured skin fibroblasts. Because the probability of recurrence is a function only of the extent of germ-line mosaicism, it is not surprising that a parent of a sporadic case could also be mosaic and in one case ($\alpha 1$ Gly-Cys 904) the mother had the features of mild OI but, or so it appeared, only on her left side (C. Constantinou, Jefferson Medical College, Philadelphia). Nor is the phenomenon restricted to the lethal form. Five recurrences due to parental mosaicism were reported in severe or mild phenotypes (G. Wallis).

The conclusion has to be that the mutations occurred not in the maturation of the germ line but very early on in the embryonic development of the mosaic parent, before the segregation of soma and germ line. This is almost bound to be a general phenomenon because there is nothing special about collagen mutations except that they are frequent and produce an easily ascertainable dominant phenotype. If the study of osteogenesis imperfecta has pushed back the origin of mutation in general to the earliest cell divisions, then perhaps marrying the girl next door was a good idea all along. □

Bryan Sykes is in the Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK.

1. Sykes, B. *Nature* **330**, 607–608 (1987).
2. Aio-Kokko, L. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6565–6568 (1990).
3. Robbeson, B.L. et al. *Molec. cell. Biol.* **10**, 84–94 (1990).

Errata

■ In "The changing shape of actin" by David J. DeRosier (*Nature* **347**, 21–22; 1990) the first full sentence on the second page was meant to read: "... the authors were able to ... locate the myosin binding site on the *small* domain of the actin monomer".

■ David Blow and Peter Brick of Imperial College London should have received full acknowledgement for kindly providing the picture of the structure of tyrosyl tRNA synthetase (**347**, 231).

■ In C. H. Llewellyn Smith's obituary of J. S. Bell (**347**, 514), the opening sentence should

have stated that Bell "also made many major contributions to accelerator physics, many-body theory — especially as applied to nuclei — quantum field theory and particle physics".

■ The final sentence of the penultimate paragraph of Fady Malik and Ron Vale's article "A new direction for kinesin" (**347**, 713–714) should have run: "Although thought to operate by different mechanisms, perhaps there are similarities between these prokaryotic and eukaryotic motors that we have yet to appreciate". □

Damming the blast

WE frequently read nowadays that some luckless car or briefcase has been destroyed by the police in a 'controlled explosion'. There is, of course, no such thing. Once you light the blue touch-paper, events run entirely out of control. Daedalus is now developing a genuine technology of controlled explosions.

In a successful explosion, he points out, each small section of exploding material transfers enough energy to the adjacent unexploded material to set it off as well. Almost all combustion generates charged molecular fragments, so some of this energy will be in the form of hot, fast-moving ions. A high electric field across an explosive might greatly increase its sensitivity and propagation rate, by driving these ions forcibly into the adjacent unreacted material. Conversely, a reversed field would drive them the wrong way, weakening the propagation of the blast. So Daedalus is wiring up sticks of dynamite to a Wimshurst machine, and studying the power and completeness of the subsequent blast as a function of the size and polarity of the applied field. Dynamite may not be ideal for this job. DREADCO's chemists are devising explosives based on caesium perchlorate, which should explode in a more ion-rich, field-sensitive manner.

This exciting research should lead to a truly controllable, field-sensitive explosive. Set off in the normal way, it will be quenched instantly by cutting off or reversing the electric field applied to it. Piezoelectric sensors attached to the object to be blown up, will integrate the acceleration or displacement caused by the blast. When exactly the right amount of force has been applied, electronic controls will reverse the field across the explosive, and just switch it off. They may even switch it on again a microsecond or so later, or modulate it into a programmed series of timed pulses. The relatively slow explosion will be completely controlled by the ultrafast electronics.

Security officers will welcome the new technique. They will be able to give a suspect package just enough shock to set off a possible hidden detonator, with no excessive damage. Tricky engineering operations like righting overturned lorries, lifting locomotives back on the rails, easing buildings apart to widen an urban bottleneck, nudging inconvenient statues of Lenin into side alleys, even blowing a safe without needless violence, all will be possible. But at this time of year, Daedalus is using the principle to invent electronically controlled fireworks. The DREADCO dot-matrix rocket, its nine parallel thrusters switching on and off rapidly under program control, will write a message in the sky, while the circular-scan catheter wheel will present a little video drama in fast-modulated fire.

David Jones

Morbillivirus in dolphins

SIR—Since late July many striped dolphins (*Stenella coeruleoalba*) in the Mediterranean Sea have died. This epizootic began on the coast of Valencia, including the Balearic Islands, and then spread to the north and south Spanish Mediterranean coasts. More than 100 dolphins have been found stranded on the Catalan coast of Spain, and the total number of adult, juvenile and newborn dolphins stranded on the Spanish Mediterranean coast is estimated to be more than 400. The epizootic is now spreading to the north African and French Mediterranean coasts.

We have taken samples from 47 dead dolphins found stranded on the coast of Catalonia. The main findings are pneumonia and encephalitis. The histopathological lesions in the animals' lungs are characteristic of a viral disease and comprise a diffuse interstitial pneumonia. We saw many multinucleated syncytia in alveolar lumina, and eosinophilic intranuclear and intracytoplasmic inclusion bodies in alveolar and bronchiolar epithelial cells (Fig. 1). Changes in the brain include neuronal degeneration and necrosis, perivascular mononuclear cell infiltration, focal malacia and syncytia formation. Many neurons and glial cells contain eosinophilic intranuclear and intracytoplasmic inclusion bodies, and there are many syncytia and severe depletion of lymphoid cells in the lymph nodes.

We examined paraffin sections of tissues from four dolphins for morbillivirus antigen by immunoperoxidase techniques. We used a polyclonal antiserum

and a monoclonal antibody against phocine distemper virus (PDV)¹ as primary antibodies, and tissues from morbillivirus-infected and healthy porpoises (*Phocoena phocoena*) as positive and negative controls, respectively. We found morbillivirus antigen in lung, brain, lymph node and biliary epithelium from all four Mediterranean dolphins using both methods. There is specific intracytoplasmic and intranuclear staining in lymphocytes and syncytia in lymph nodes (Fig. 2).

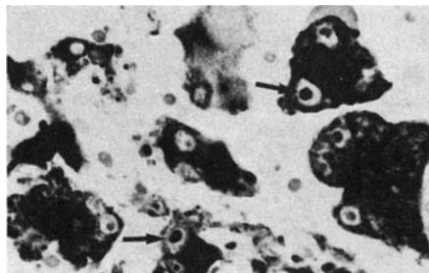


FIG. 2. Immunoperoxidase staining of morbillivirus antigen in cytoplasm and nuclei (arrows) of syncytia in lymph node using a monoclonal antibody against PDV. $\times 510$.

Antigen also occurs in bronchial and bronchiolar epithelial cells, pneumocytes and syncytia in the lungs, and in neurons and glial cells in the brain. Electron microscopic examination of lung tissue from affected dolphins reveals viral particles morphologically consistent with a paramyxovirus in the cytoplasm of lung epithelial cells and syncytia.

We attempted to isolate viruses from the tissues of three dolphins. A cytopathic virus was cultivated from brain, lung and mediastinal lymph nodes of all three animals. Bovine and canine kidney cell lines infected with this virus contain eosinophilic intracytoplasmic and occasionally intranuclear inclusions, and give a positive immunoperoxidase reaction when incubated with hyperimmune serum against CDV from a dog, confirming that the isolated virus is a morbillivirus.

The microscopic changes in these dolphins are similar to those found in common seals (*Phoca vitulina*)^{1,2} and porpoises³ during the northern European seal morbillivirus epizootic in 1988, and in dogs with canine distemper⁴. Our demonstration of morbillivirus antigen in diseased tissues provides conclusive evidence for a primary aetiological role of PDV (or a closely related morbillivirus) in development of the lesions.

This is the first report of morbillivirus infection in dolphins. Further studies to characterize the isolated virus and to compare it with isolates of seal² and porpoise³ morbillivirus are in progress which should determine if there is a link between the 1988 and 1990 epizootics. Further, the presence of morbillivirus

infection in striped dolphins may be a serious threat to the survival of the small population of Mediterranean monk seals (*Monachus monachus*), already considered an endangered species.

M. DOMINGO
L. FERRER
M. PUMAROLA
A. MARCO

Department of Veterinary Pathology,
Autonomic University of Barcelona,
08193 Bellaterra, Barcelona, Spain

J. PLANA

Sobrino Laboratories,
17813 Vall de Bianua,
Olot, Girona, Spain

S. KENNEDY*
M. MCALISKEY

Department of Agriculture for Northern
Ireland,
Veterinary Research Laboratories,
Stormont, Belfast BT4 3SD,
Northern Ireland

B.K. RIMA

School of Biology & Biochemistry,
The Queen's University of Belfast,
Belfast BT7 1NN,
Northern Ireland

1. Kennedy, S. *et al.* *Nature* **335**, 404 (1988).
2. Kennedy, S. *et al.* *Vet. Path.* **26**, 97 (1989).
3. Kennedy, S. *et al.* *Nature* **336**, 21 (1988).
4. Dungworth, D.L. in *Pathology of Domestic Animals* Vol. 2 (eds Jubb, J.V.F., Kennedy, P.C. & Palmer, N.) Ch. 6 (Academic, London, 1985).
5. Cosby, S.L. *et al.* *Nature* **336**, 115 (1988).

* To whom correspondence should be addressed.

Fallout in snow

SIR—Dibb *et al.* in their recent Scientific Correspondence¹ questioned the origin of a caesium-137 peak measured in a snow sample collected near the South Pole and corresponding to deposition during the summer of 1987–88. Because they found no ¹³⁴Cs, characteristic of Chernobyl fallout, in their samples they were not sure that the Chernobyl accident could explain the observed radioactivity.

I believe that the radioactive contamination is certainly due to the Chernobyl accident. Indeed, my colleagues and I have detected ¹³⁴Cs in atmospheric dust samples taken from two stations in the Southern Hemisphere, Tahiti and La Réunion, during January 1987. The levels of radioactivity were low (a few microbecquerels per m³) but the ¹³⁴Cs/¹³⁷Cs ratio was similar to the ratio observed in European countries affected by the accident. Furthermore, observations made in Poland by Jaworowski *et al.*² indicate that aerosols emitted during the accident were present in stratospheric air.

J.C. PHILIPPOT

Laboratoire de Radioactivité de l'Air,
Bât. 501 – Bois des Rames,
F-91400 Orsay, France

1. Dibb, J.E. *et al.* *Nature* **345**, 25 (1990).
2. Jaworowski *et al.* *J. environ. Radioact.* **6**, 145–150 (1988).

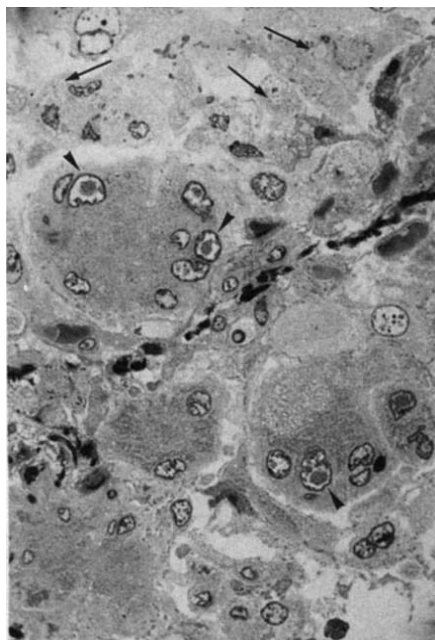


FIG. 1. Multinucleated syncytia formed by fusion of morbillivirus-infected lung epithelial cells. Intranuclear (arrowheads) and intracytoplasmic (arrows) inclusion bodies are seen. Toluidine blue stain. $\times 700$.

Sulphate aerosol and climate

SIR—Wigley has suggested in Scientific Correspondence¹ that the Northern Hemisphere may be warming more slowly than the Southern Hemisphere, possibly due to an anthropogenic increase of cloud condensation nuclei of about 20% in the Northern Hemisphere. We wish to propose an alternative yet complementary explanation for the suggested difference — that backscatter of solar radiation by non-cloud, anthropogenic SO_4^{2-} aerosol particles in cloud-free air significantly reduces the incoming solar radiation over much of the Northern Hemisphere. A simple box-model calculation illustrates the expected magnitude of the mean column burden of anthropogenic sulphate, $B_{\text{SO}_4^{2-}}$:

$$B_{\text{SO}_4^{2-}} = \frac{F_{\text{SO}_4^{2-}} \tau_{\text{SO}_4^{2-}}}{A} \sim \frac{(3.3 \times 10^6 \text{ g s}^{-1}) (5 \times 10^5 \text{ s})}{2.5 \times 10^{14} \text{ m}^2} \\ \sim 6.6 \times 10^{-3} \text{ g m}^{-2}$$

where $F_{\text{SO}_4^{2-}}$ is the average flux of this SO_4^{2-} through the atmosphere in the Northern Hemisphere (equivalent to about half of 70 Tg yr^{-1} of sulphur emitted as SO_2); $\tau_{\text{SO}_4^{2-}}$ is the mean sulphate aerosol particle lifetime (about 6 days) and A is the area of the Northern Hemisphere. We assume that all anthropogenic SO_4^{2-} originates and stays in the Northern Hemisphere. An empirical optical scattering efficiency, α ($10 \text{ m}^2 \text{ g}^{-1}$, ref. 2), then yields an estimate of optical depth, $\delta_{\text{SO}_4^{2-}}$, for solar wavelengths due to anthropogenic SO_4^{2-} :

$$\delta_{\text{SO}_4^{2-}} = \alpha B_{\text{SO}_4^{2-}} \sim 0.066$$

Finally, an empirical backscatter fraction, β (0.15, ref. 3), and estimated cloud fraction, f (~ 0.6), allow for estimation of the energy lost from the Earth's surface, L (we disregard the effect of aerosols above cloudy areas):

$$L \sim (1 - f) \beta_{\text{SO}_4^{2-}} (2\delta_{\text{SO}_4^{2-}}) \sim 0.8\%$$

where the factor of two is the secant of solar zenith angle averaged over the sunlit hemisphere.

If the solar irradiance incident on this low-altitude aerosol is about 240 W m^{-2} , irradiance (direct plus diffuse solar radiation) is lost at a rate of about 2 W m^{-2} . The solar energy lost from the Earth-atmosphere system would be somewhat less — approximately 1.6 W m^{-2} — due to the albedo (reflectivity) of the underlying surface. This radiative forcing is comparable in magnitude but opposite in sign to the current forcing by anthropogenic CO_2 of about 1.5 W m^{-2} (ref. 4). Hence, backscatter of solar radiation by SO_4^{2-} aerosol cannot be ignored and could account for some or all of the difference in temperature trends between the two hemispheres¹. In addition to this direct effect, anthropogenic SO_4^{2-} may have an indirect

effect by acting as cloud condensation nuclei which can influence cloud albedo⁵. However, this cannot be quantified because there is no known/accepted relationship of SO_4^{2-} mass concentration to the number concentration of cloud condensation nuclei.

Refinements are possible in the calculation of the non-cloud aerosol radiative forcing. Sulphur chemistry has been included in a three-dimensional global model⁶ to estimate the geographical dependence of $\delta_{\text{SO}_4^{2-}}$. We already know that anthropogenic SO_4^{2-} concentrations and loss of solar irradiance are highest in industrial regions, low over the North Pacific and essentially absent in the Southern Hemisphere. The radiative transfer calculation also can be refined with the addition of well-developed models, including a full treatment of the angular scattering characteristics, seasonality, as well as the geographical distribution of cloud fraction, vertical distribution of aerosol and surface albedo.

Uncertainties will remain even when such a coupled chemical/radiative transfer model is done, because the geographical nonuniformity and asymmetry of the Northern and Southern Hemisphere

forcing may cause shifts in meteorological parameters unlike those due to the geographically uniform forcing by greenhouse gases. It does seem unlikely, however, that the magnitude of irradiance loss estimated above will change radically in a refined model calculation. Thus, the effect of anthropogenic aerosol has to be taken into account in any assessment of climate change.

R. J. CHARLSON

Department of Atmospheric Sciences,
University of Washington,
Seattle,
Washington 98195,
USA

J. LANGNER

H. RODHE

Department of Meteorology,
Stockholm University,
S 106 91 Stockholm,
Sweden

1. Wigley, T.M.L. *Nature* **339**, 365–367 (1989).
2. Waggoner, A.P. *et al.* *Nature* **261**, 120 (1976).
3. Vanderpol, A.H. thesis (University of Washington, Seattle, 1975).
4. Intergovernmental Panel on Climate Change Report of Working Group 1 (World Meteorological Organization and United Nations Environmental Programme, Cambridge University Press, 1990).
5. Twomey, S. *J. Atmos. Sci.* **34**, 1149–1152 (1977).
6. Langner, J. & Rodhe, H. *Proc. Int. Conf. Global Regional Envir. Atmos. Chem.* (eds Newman, L., Wang, W. & Kiang, C.S.) 106–117 (US Dept. of Energy, Washington, DC, 1990).

Differences in hair and fleece loss

SIR—Paul Bowden in his excellent News and Views article¹ on gene expression and hair loss made an oversight in his passing comment that sheep have a mosaic pattern of wool replacement as is the case for hair growth in man. Most mammals have a seasonal pattern of coat replacement in which the skin follicles are dormant during the winter (telogen) and active during the summer (anagen) after a spring moult. This annual cycle of growth is controlled by changes in day length and is universal among the hoofed animals, including the wild ancestor of domestic sheep².

Shortening days in autumn trigger changes called catagen, leading to the winter resting period (telogen), whereas lengthening days in the spring stimulate re-growth (anagen), leading to the loss of the old coat. This re-growth before loss of the old coat is a mechanism, like that found in mosaic replacement, which ensures that the animal is never naked as a result of the moult.

The shorter cycles of hair replacement in mice discussed by Bowden, which occur in waves across the body, and the mosaic replacement in man (and guinea pigs) are highly specialized. The lack of a seasonal link in the guinea pig can be interpreted as due to evolution near the equator, where major seasonal climate changes necessitating summer and winter coats do not

occur and there is little accompanying variation in day length.

From the Iron Age onwards, sheep were selected for continuous wool growth so that wool could be shorn and not lost during the moult³, and an evolutionary sequence can be traced from surviving primitive breeds, which have a complete moult, to modern breeds, which have little tendency to shed their wool⁴. Bowden probably had the Merino breed of sheep in mind when he said that man and sheep have a mosaic system of active hair replacement because the Merino more than any other breed has virtually continuous wool growth, mimicking a true mosaic system of wool replacement. But an incidence of dormant follicles of 4 per cent in the Merino, at the time of the year when primitive sheep would also have dormant follicles, does not constitute mosaic replacement⁵. That level of dormancy means that Merino wool fibres are replaced only once every 8 years, and as this is longer than the life of the animal it is not too much of an exaggeration to say that Merino sheep in truth grow wool continuously.

M. L. RYDER

4 Osprey Close,
Southampton SO1 8EX, UK

1. Bowden, P. *Nature* **346**, 315 (1990).
2. Ryder, M.L. & Stephenson, S.K. *Wool Growth* (Academic, London, 1968).
3. Ryder, M.L. *Sheep and Man* (Duckworth, London, 1983).
4. Ryder, M.L. *Int. J. Chronobiol.* **5**, 369–394 (1978).
5. Ryder, M.L. *Aust. J. agric. Res.* **18**, 683–88 (1967).

A surfeit of models

Edward S. Reed

Modelling The Mind. Eds K. A. Mohyeldin Said, W. H. Newton-Smith, R. Viale and K. V. Wilkes. Clarendon: 1990. Pp. 216. £25, \$49.95.

The Computing Neuron. Eds Richard Durbin, Christopher Miall and Graeme Mitchison. Addison Wesley: 1990. Pp. 417. £25.95.

IN HIS contribution to *Modelling the Mind*, P.N. Johnson-Laird, arguably the doyen of experimental studies of reasoning, warns that "people . . . usually hold that a conclusion is beyond reasonable doubt when they have exhausted those possibilities that are obvious (to them). It is easy to show that this judgement is often premature." This kind of 'argument from exclusion' is a common danger in scientific as well as in ordinary thinking. A case in point are the two volumes under review, which pose for readers the much-discussed choice of mental versus neural modelling. For three decades or so, cognitive scientists have been pursuing the implications of the assumption that the mind is analogous to a (digital) computer. Despite some important successes, this point of view is at present under strenuous attack, even being supplanted in certain research areas by 'connectionist' approaches which assume that the brain is a (massively parallel) computer. *Modelling the Mind* focuses on the mind-computer analogy, although most of the chapters at least acknowledge the potential importance of connectionism. By contrast, throughout *The Computing Neuron* it is simply assumed that the brain is a computational network.

Reading the two in tandem convinces me that the choice between mind-as-computer and brain-as-computational network in no way covers the possibilities, and that what is needed in cognitive science are theories that use the great wealth of available psychological and neurophysiological data to explore alternatives to computationalism. One can agree with the connectionists that the mind-computer analogy has outlived its usefulness without jumping to the conclusion that brains are computers. After all, these analogies are useful only insofar as they enable us better to understand psychological or neural processes, and it is easy to feel that we have come to the point of limiting returns in both these areas.

Biological honesty

One of the goals of *The Computing Neuron* is to fashion a connectionism that is biologically honest. Each editor contributes a useful introduction: Durbin on the relation of hypothetical neural units to real neurons, Miall on variation in neuronal properties, and Mitchison on learning networks. Unfortunately, none

of them and few of the contributors succeed in connecting the very real powers of their neural nets to principled accounts of biological behavioural function. They seem to think that models can float free of either biological or psychological theories so that, if the hypothetical units of their nets have some of the properties found in some neurons, then this counts as biological realism. The authors show a lack of concern for using models to test behavioural theories and hypotheses. Time and again the reader will wonder what the model being developed is a model of and — more important — why this particular capacity was singled out for modelling.

The more interesting models discussed in the book all derive from strong theories, precisely because these theories provide powerful constraints on the testing and evaluation of models in their domain. Koch and his colleagues, for example, develop a neural model for the detection of optic flow which, thanks to the work of perception theorist J. J. Gibson, we know is a fundamentally important visual process. There are several algorithms available that can accomplish the derivation of flow from image sequences, and Koch's group takes one such algorithm and integrates it nicely with the wealth of neurophysiological data about primate visual cortex. In particular, this group pays careful attention to the kind of 'population coding' by which features of stimulation seem to be extracted in the sensory cortices. This is a model which can be checked against robust data from research on visual perception and neurophysiology. Yet even in this case, there is a remarkable lack of concern for some basic issues in functional biology. For example, despite these authors' care to match the operating principles of the visual central nervous system, they use a model of retinal image optics that is totally at variance with the facts of the ever-exploring primate visual system. The book includes a brief chapter on the salamander retina, but no useful information on ecological optics or on the comparative morphology and physiology of ocular image formation, without which it will be impossible to produce biologically realistic models of visual perception.

Evolved systems do not exhibit rational design properties. One of Darwin's most important contributions was to emphasize how widespread are dysteleology and

historical accident in nature. Only one of the chapters in *The Computing Neuron*, Robertson's fine brief review of the neuronal circuitry underlying locust flight, emphasizes this darwinian point. Yet if central nervous systems have been evolved by natural selection, then, as Robertson aptly puts it, "many details of [their] operation may be suboptimal or baroque", historically derived contraptions. Hence, the precise modelling of a given behaviour is of dubious value because a great deal of the detail will be more relevant to phylogenetic idiosyncrasy than to neurobehavioural principle. Robertson drives this point home with a diagram of the great variability found in an identified neuron in five individual locusts. These insects' neural nets are composed of individual cells of such variety that prediction of circuit properties of these 'simple' nervous systems has not been possible. This failure of prediction is worth remembering when evaluating any of the successful but *post hoc* neural-network models.

Local variability

The state of the art in neuronal modelling is thus highly paradoxical: network models of processes as complex as reading exist side by side with serious failures to elucidate principles even in simple invertebrate systems where many neurons and circuits have been identified and characterized. Until biological realism means more than having hypothetical units copy features of neurons, such paradoxes will be commonplace. Neural modellers need to address such questions as why many features of neural units and connections can vary between individuals without compromising the species' neuronal functioning. The local variability of nervous systems is not noise to be added to one's model of nervous systems but a biological principle that requires explanation.

Modelling the Mind, by contrast, is concerned with psychological function at a fairly high level. Considering the scope and ambitions of some modellers, it is surprising how infrequently the question "what is it we should be modelling when we model the mind?" is asked. Only Wilkes raises it in this collection. She uses the example of 'memory', pointing out that this apparently simple mental process is now widely seen as being a conglomeration of different capacities (short or long term, semantic or episodic, item or autobiographical and so on). "We cannot say that we have another instance of 'the same' capacity in a computer until we know what it is that we have got." In general, the psychological processes which we find most intriguing — imagination, thinking, perceiving and so on — are described by words for which the referents are not clear. As Davidson explains in his marvellous critique of 'Turing's

test', we know these capacities when we see them, but we cannot make explicit what it is we know. This interpersonal recognition is yet another of these marvellous psychological skills. If Davidson is right, our intuitions, which emerged during the evolution of ourselves as social primates, may work for creatures like ourselves but are suspect when applied to different entities like computer models of the mind. The moral of this is not despair, but the need to develop hypotheses that are precise enough to be tested about these intriguing psychological processes. This sort of nonglamorous hypothesis-testing has been the mainstay of not a few cognitive psychologists (such as the work on reasoning and thinking reviewed here by Johnson-Laird and Evans) but it seems that neuroscientists and artificial intelligence researchers are reluctant to take advantage of the gains of their psychologist colleagues. And some philosophers seem wilfully ignorant, as witness the nonsensical chapter by Boden in which she uses Spinoza as the model of scientific psychology preceding artificial intelligence, ignoring three centuries of research.

Noble's intriguing essay comparing Hodgkin-Huxley type models of the nerve impulse to computer models of the mind suggests that without rigorous theories of psychological functioning we may not even be able to interpret the models of lower level processes we already have. The Hodgkin-Huxley model would certainly be of far less interest or use if we had not already a considerable body of theory and data about the higher-level, integrative functioning of the nervous system, within which context the model could be interpreted.

There is no easy short cut to a science of the mind, computer technology and its advances notwithstanding. What is needed in cognitive science is good thinking tempered by careful research unhindered by restrictive models and analogies. After this, modelling might help to clarify some issues. For Johnson-Laird, everyday mental models of the world are acts of constructive imagination, but he emphasizes that it is imagination that is central in these models, not inference. Similarly in science, much of the effort of thought lies in the theoretical imagination, with the kind of inferencing characteristic of formal models playing a secondary role. Cognitive science has had far more than its share of modellers, but precious few thinkers and theorists. If ever there were a task that required good imagination intermingled with hard thought, it is the job of understanding the human brain and mind. □

Edward S. Reed is at Drexel University, Philadelphia, Pennsylvania 19104 and the Drucker Brain Injury Center, Moss Rehabilitation Hospital, Philadelphia, Pennsylvania 19141, USA.

Star exhibits

A.C. Fabian

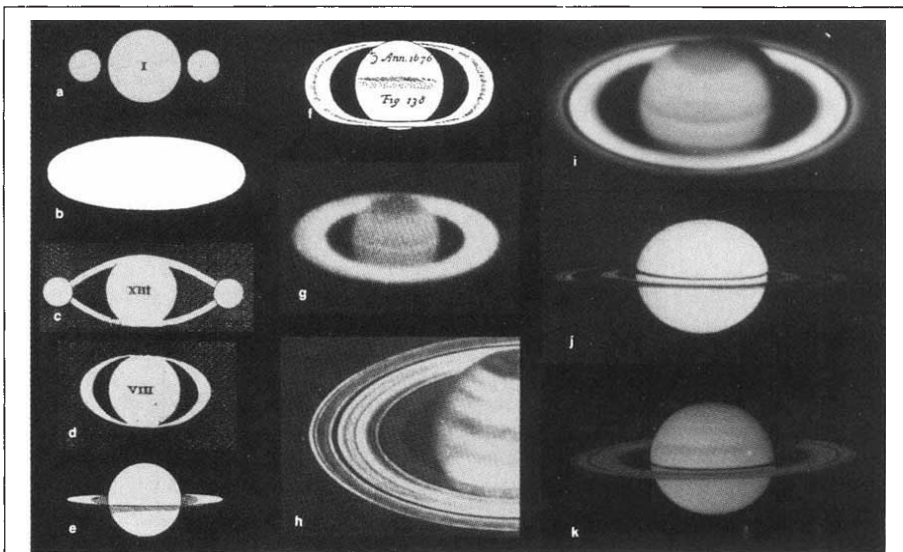
The Astronomer's Universe. Stars, Galaxies and the Cosmos. By Herbert Friedman. Norton: 1990. Pp. 359. \$24.95, £19.95.

WE ARE taken on a tour of a museum by an amiable guide. In this cabinet we have a red giant, over there we have the Hale telescope and just come and look at the solar flare round the corner. Here we have the eminent Soviet astrophysicist about whom it was said that "fifty per cent of

of a nonsolar X-ray background.

This early work culminated in the *Vanguard* satellite programme, which had a disastrous beginning. Friedman's first *Vanguard* satellite was submersible instead of orbital. His second one was swamped by radiation from the then recently discovered Van Allen belts. Nevertheless, persistence won through and results began to flow in.

One curious aside, made in 1959, notes that the steady-state theory of the Universe, as calculated by Fred Hoyle at that time, predicted a background X-ray flux 100 times higher than Friedman's data allowed. This calculation was presumably forgotten about in 1962 when Hoyle



Ringing the changes — Our view of Saturn's rings has improved over the centuries, from Galileo's drawing in 1610 (a) through various sketches and photos to Voyager 1's image in 1980 (k). *The New Solar System*, eds J. K. Beatty and A. Chaikin, describes the Solar System as we know it today, thanks to modern astronomy and robotic probes. Published by Cambridge University Press, price is £25, \$39.95 (hbk); £13.95, \$24.95 (pbk). □

what he does is brilliant but no one can tell which fifty per cent it is", and an extract of a letter from Thomas Edison's associate with his proposal to detect magnetic disturbances from the Sun by using a giant mass of iron ore with wire wound around it — in 1890! This is *The Astronomer's Universe*, and Herbert Friedman of the Washington Naval Research Laboratory is the guide.

The description on the exhibits are rather dull, probably through over-use, although the exhibits are interesting enough. When the guide pauses for an anecdote, particularly if related to his own research experiences in the early days of rocket astronomy, the story takes off and the excitement and the adventure of frontier research shines through. Friedman and the Naval Research Laboratory were involved in early experiments with captured V-2 rockets, flying spectrometers, photographic plates and photon counters from White Sands to altitudes above 100 km in the late 40s and early 50s. They discovered ultraviolet and X-ray emission from the Sun and some hints

published a paper claiming that the X-ray background just discovered by Giacconi and his co-workers supported a steady-state cosmology.

Perhaps the most currently relevant aspect of these tales is the trial-and-error and try-and-try-again nature of the research. Generally it was successful and when it wasn't, it was usually the fault of the apparatus or the rocket. Serendipity played, and still plays, an enormous role. Most cosmic phenomena have been found by chance. Observations lead the subject and theory provides an essential later synthesis. How this era of discovery would have proceeded if it had all gone through the present peer-review system, which is usually overloaded by more than three good proposals to each one accepted, does not bear thinking about. Discoveries often come from an advanced form of playing and playing should not be made accountable. What does bear thinking about is how our present systems stifle enthusiastic trial-and-error research.

I found the tales of the early rocket days made interesting reading and wished that

they were a larger fraction of the book. But then I am not the 'layperson' for whom it was written or selected by the Commonwealth Fund book programme. I can certainly recommend it for someone wanting a readable overview of modern astronomy, livened up with anecdotes and without the myriad glossy pictures common in many such books. □

A. C. Fabian is at the Institute of Astronomy, University of Cambridge, The Observatories, Madingley Road, Cambridge CB3 0HA, UK.

Rodent preserve

Donald K. Grayson

Packrat Middens. The Last 40,000 Years of Biotic Change. Edited by J. L. Betancourt, T. R. Van Devender and P. S. Martin. University of Arizona Press: 1990. Pp. 467. \$55.

FROM Nicaragua to northern Canada, packrats (*Neotoma*, Muridae) accumulate plant material and other items in and near their dens. In arid settings, the impregnation of these accumulations by packrat urine produces dark, indurated middens. During the 1960s, P. V. Wells and C. Jorgensen demonstrated the existence of plant-rich middens of Pleistocene age in southern Nevada, and in so doing introduced a formidable tool that could be used to probe the palaeoecology of the arid west: a source of late Pleistocene and Holocene plant fossils that could be accurately dated.

More than a thousand packrat middens from arid western North America have now been analysed and dated, much of that work done by students of Paul S. Martin of the University of Arizona's Desert Laboratory. *Packrat Middens: The Last 40,000 Years of Biotic Change* summarizes what has been learned, where the problems may lie, and what the future may bring. With the exception of Wells himself, all the major names in this small discipline are represented here.

Packrat Middens begins with an assessment of the methodological difficulties that these middens present, emphasizing that different species of packrats have markedly different diets and that all packrats collect biased samples of plants from around their dens. Indeed, reoccupation of identical midden sites by different species of packrats can result in the accumulation of very different plant assemblages. There is bias in the accumulated sample as well: most middens date to the late Pleistocene or late Holocene, with a conspicuous lack of examples from the mid-Holocene. In addition, our current midden sample is not randomly distributed across slope aspects or rock substrates. Given that plants respond sensitively to

these variables, the potentials for bias are clear.

It is a necessary hallmark of palaeoecologists that they worry about severe biases in their data at the same time as they go about their business. This book is no exception, presenting superb midden-based syntheses of the past vegetation of the Chihuahuan, Sonoran, Mojave and Great Basin deserts, as well as of the Colorado plateau and the Grand Canyon. Readers may be astonished to learn that the lower elevation of North America's warm deserts were covered by woodlands during the end of the Pleistocene and into the early Holocene, and that the vast ponderosa pine forests of the Colorado plateau appear to be a Holocene phenomenon.

The dynamic plant histories are fascinating, but the themes and debates that tie these syntheses together give them global value. The authors are unanimous in emphasizing the individualistic responses of plant species to climate change; there are no plant communities marching across space to be found here. Less unanimity is to be found over the issue of vegetational inertia, the argument that plants may respond very slowly to major aspects of climate change. Even more debate greets the results of the climate modellers. W. G. Spaulding argues that there were strengthened summer monsoons in the Mojave desert in the early Holocene, as is required by the COHMAP model, but T. R. Van Devender amasses equally convincing evidence that the Holocene monsoonal maximum in the Sonoran desert occurred in the middle, not the early, Holocene. For the Great Basin desert to the north, R. S. Thompson notes that the early Holocene thermal maximum called for by the COHMAP model simply did not happen.

Potential directions for the future are also provided. A series of analyses focuses on subsets of materials from packrat middens—grasses, mammals, insects and the deuterium content of plant cellulose. These analyses are weaker than the general syntheses: samples are smaller, the causes of variations in deuterium content poorly understood, and modern arthropod faunas and morphology insufficiently known to tap fully the information provided by their fossil counterparts. The potential, however, is clear. Grasses identified to the species level allow Van Devender and his colleagues to question the argument that parts of the Sonoran desert supported a climax grassland before the introduction of domestic herbivores; Van Devender and G. L. Bradley show that during the past 15,000 years, arthropod assemblages seem to have been far more stable than plant assemblages in southwestern Arizona.

These studies demonstrate that increasingly detailed analyses of packrat middens

are likely to bring spectacular results. Similar opportunities exist in other parts of the world. Hyraxes (Procaviidae) and dassie rats (Petruridae) create middens in Africa and the Middle East; stick-nest rats (*Leporillus*) created them in arid Australia. An ancient world clearly awaits discovery.

Packrat Middens provides splendid accounts of the dynamic vegetational histories of North American deserts. Like other fossil-based assessments of Quaternary plant and animal histories, this book shows that modern biotic distributions cannot be understood without deep historical data. Let us hope that the Desert Laboratory, the intellectual source of this remarkable volume, will continue to thrive as long as there are packrats gathering plants in the American deserts. □

Donald K. Grayson is in the Department of Anthropology and the Quaternary Research Center, University of Washington, Seattle, Washington 98195, USA.

Problems with diversity

Alan Brafeld

Invertebrates. By R. C. Brusca and G. J. Brusca. Sinauer: 1990. Pp. 922. £27.95, \$47.50.

THERE are some problems in zoology which defy solution but which will not go away. One is how to make sense of invertebrate diversity. There are more than 30 invertebrate phyla, each built on a unique ground-plan. How can they (and the subgroups within them) be meaningfully related, particularly with so little fossil evidence? Most invertebrate textbooks simply deal with each group in turn, but a few consider 'systems' on a comparative basis — feeding, respiration, movement and so on — a functional rather than a taxonomic view. Two years ago I wrote here of how successfully R. S. K. Barnes *et al.* in *The Invertebrates: A New Synthesis* had attempted both approaches, by first surveying the groups and then reviewing comparative functional biology. Brusca and Brusca have here tried to fuse the two. In the opening chapters they develop three themes: functional body architecture, patterns of development and life histories, and evolution and phylogeny. They then embark on the familiar tour through the phyla, but keep these themes well to the fore.

Unfortunately, the authors are not completely successful with their dual aims. They provide an immense amount of detailed information about all the groups, including the recently discovered ones, and so this is an invaluable and

up-to-date reference work. But the book almost sinks under its own weight (nearly six pounds!). Will students be able to follow the concepts through the plethora of detail? It was the authors' intention to produce a work of about 400 pages. "Obviously, we were unable to do that." Obviously, but what a pity. Pruning of much fringe material would have let the linking ideas emerge more vividly.

The cladistic approach to the treatment of relationships is carefully explained and is welcome. Page references in the index in bold type show where a definition of the term is to be found, so combining an index with a glossary — another good idea. The illustrations (mainly by Nancy Haver) are excellent and plentiful and a wide selection of references is given at the end of each chapter.

With perhaps even more information than in R.D. Barnes' *Invertebrate Zoology*, and with the 'threads of continuity' so well thought out, *Invertebrates* deserves to succeed; but to get the most out of it in terms of unifying concepts rather than detailed facts, students will need to read it with great care and selectivity. □

Alan Brafield is in the Division of Biosphere Sciences, King's College London, Campden Hill Road, London W8 7AH, UK.

Forest matters

Kenneth Mellanby

Global Forest Resources. By Alexander S. Mather. *Belhaven*: 1990. Pp. 341. £40.

Australian Tropical Rainforests. Edited by L J Webb & J Kikkawa. *CSIRO Publications*: 1990. Pp. 185. \$55.

THE importance of the world's forests is becoming increasingly understood. They produce timber for buildings and furniture; fuelwood is the main energy source for much of the world's population; they provide food, both animal and vegetable, and an increasing number of drugs and other chemicals. Conservationists stress the importance of the vast diversity of species, many of which have not yet been identified, to be found in tropical forests. Woodland is increasingly used for recreation in developed countries.

The destruction of forests, liberating carbon dioxide when the wood is burned, and reducing the uptake of that gas by growing trees, contributes to the build-up of the so-called greenhouse gases. Although forest clearing has produced much of the world's farmland, so allowing food production to keep up with population growth, recent felling in the tropics, like some deforestation in North America in the nineteenth century, has resulted in erosion, soil degradation and flooding. In



Forest tails — These spider monkeys feature in *Neotropical Rainforest Mammals: A Field Guide* by L. H. Emmons and F. Feer, which describes the mammals of the rainforests, their natural history and conservation status. Published by The University of Chicago Press, price is \$51.75, £35.95 (hbk); \$22.95, £15.95 (pbk). □

his book *The Ages of Gaia*, James Lovelock has suggested that tropical forests affect global climate, and that their loss could have results even more serious than those foretold for the greenhouse effect.

Alexander Mather does not underestimate the losses that have occurred, from the Neolithic period onwards. The effects of forest clearances round the Mediterranean in the time of the Roman empire have had permanent results on many countries, once fertile and now semi-desert. In Britain, forest areas in 55 BC when Julius Caesar invaded the country were smaller than today. Notwithstanding accelerated deforestation in many areas, Mather records that "as much as 80 per cent of the [world's] pre-agricultural forest survives". This situation is not one to encourage complacency, but it does suggest that increased efforts to preserve our remaining forests are still very much worth while.

Mather shows how, in developing countries, after a period of destruction, there is often a 'steady state' where new planting is at least equal to the areas being cleared. But much of this planting is of quick-growing exotic species. Thus Australian Eucalyptus species are the commonest trees in large parts of Africa, South America and China, and American Sitka Spruce has contributed largely to new planting in Europe. These new plantations are more satisfactory to foresters than to ecologists, who deplore the loss of ancient woodlands and the virtual disappearance of virgin forest from developed countries.

In *Global Forest Resources*, Mather gives a succinct account of the world situation. He includes a brief history of forests and their exploitation, and describes in some detail today's position. The possible exploitation of forests on a sustainable

basis is the main theme throughout the book. As already mentioned, the use of fuelwood throughout the developing world is a major international problem. We all have had our feeling harrowed by television programmes showing women spending many hours collecting and carrying the twigs that are their only fuel. Planting in areas of shortage are not always successful in the most affected areas, where drought is frequent. However, in other parts of the world 'biomass' is increasingly realized to be an important source of energy, and as it recycles the carbon dioxide absorbed by fuel trees it makes no net contribution to greenhouse gasses.

The Australian Tropical Rainforests of Queensland, described by 22 experienced scientists in the volume edited by L. J. Webb and J. Kikkawa, can be considered a special case illustrating some of Mather's points. The main purpose of this book, however, is to draw attention to the value of the forests described. North Queensland's rainforest is "the only part of mainland Australia that has persisted continuously as wet tropics for the last 65 million years (since the Cretaceous period)". Its interest value economically, scenically and even spiritually is the theme of the writers. They are also concerned with the dilemma of whether to permit selective logging of valuable trees for timber. This has been suggested as the only economic alternative to total clearance. Most of the authors consider that any exploitation would be damaging, and that these unique forests are so important that they should be strictly preserved for posterity. □

Kenneth Mellanby is at 38 Warkworth Street, Cambridge CB1 1EG, UK and was formerly director of Monk's Wood Experimental Station.

Optimality theory in evolutionary biology

G. A. Parker & J. Maynard Smith

Optimization models help us to test our insight into the biological constraints that influence the outcome of evolution. They serve to improve our understanding about adaptations, rather than to demonstrate that natural selection produces optimal solutions.

IN recent years, optimization theory and game theory have been widely used, particularly by field biologists, to analyse evolutionary adaptation¹⁻⁷. During the same period, originating with a classic paper by Gould and Lewontin⁸, there has been continuing criticism of the optimization approach. This criticism seems to arise from the idea that those who adopt the approach assume either that animals and plants are optimally adapted, or that they are trying to prove that they are so. Hence any demonstration of the role of chance events—for example, that much molecular variation is selectively neutral, or that unpredictable events have had a major effect on evolution—has been seen as undermining the optimization approach. If, by this review, we could lay rest to the idea that the application of optimization theory requires either that we assume, or that we attempt to prove, that organisms are optimal, we would be well satisfied.

It is true that the optimization approach starts from the idea, already familiar to Darwin, Wallace and Weismann in the last century, that adaptation is a pervasive feature of living organisms, and that it is to be explained by natural selection. It is not our aim to add to this the claim that adaptation is perfect. Rather, the aim is to understand specific examples of adaptation, in terms of the selective forces and the historical and developmental constraints operating. This requires that we have an explicit model, in each specific case, that tells us what to expect from a given set of assumptions. The predictions are an inevitable consequence of the assumptions. A model cannot then be 'wrong' (unless analysed incorrectly), but it can certainly be inappropriate if it is based on assumptions that are not well founded.

We distinguish between general models and specific models⁹, though in reality they form part of a continuum. General models have a heuristic function; they give qualitative insights into the range and forms of solution for some common biological problem. The parameters used may be difficult to measure biologically, because the main aim is to make the analysis and conclusions as simple and direct as possible. Specific models are designed to be applied quantitatively to particular species, and include parameters that can readily be measured. They are often modified (and more complex) versions of some general model, devised specifically for comparison with a particular set of observations. If the predictions of the model match the biological observations, we may hope that we have made correct assumptions about the nature of adaptations.

Evolutionary optimization models

The optimality approach involves the construction of a model about adaptation. First, it is necessary to ask an explicit biological question, which may be general (for example, why is the sex ratio often unity?¹⁰⁻¹²) or specific (for example, why do dungflies copulate for 36 minutes?¹³). If we do not know that the sex ratio is often unity, or that dungflies copulate for 36 minutes, then the equivalent questions might become: what sex ratio is the most likely to result from natural selection? and how long should dungflies copulate as a result of male-male competition? Either way, the question is assumed to have an adaptive answer, otherwise we cannot proceed to establish whether a given adaptive process can generate the correct solution.

Next, a range of alternative actions or 'strategies' relating to

the question is defined. For the question about sex ratios, the obvious strategy set (the range of phenotypic variants) to consider is all points in the continuum from producing only male offspring to producing only female offspring. But strategy sets need not be continuous: many models involve discrete strategies. Thus, for a bird's choice of where to nest, alternative strategies might include nesting in a tree or nesting on the ground.

The strategy set simply specifies the plausible alternatives, given what we consider it possible for evolution to achieve. Often, as in the case of sex ratios, there is a simple and obvious strategy set that logically covers all the possibilities. But in other cases it is necessary to rely simply on biological intuition. Some feel for candidate strategies can often be obtained from the existing range of variation. Typical biological constraints usually define some obvious boundary conditions for the strategy set, but strategic possibilities that have never in fact been observed are included unless there are reasons of this kind for leaving them out.

In the construction of the model an assumption must be made about what is being maximized, and some measure of darwinian fitness is usually used. The simplest direct criterion is the expected lifetime number of surviving offspring produced by an individual pursuing a given strategy ('individual fitness'), defined in units of generation time (which may vary with phenotype). Many life-history theorists prefer to use the more cumbersome rate of increase per individual (r , the solution of the Euler-Lotka equation¹⁴). But in most circumstances, the relative reproductive outputs of different strategies (which equate roughly with the selection coefficients used in classical population genetics) will suffice. When interacting individuals are relatives, however, it is more satisfactory to use Hamilton's 'inclusive fitness'¹⁵, which sums, for an allele prescribing a given strategy, the aggregate consequences for that allele (inclusive of the same allele carried in relatives). Other non-darwinian criteria such as group selection (the notion that selection minimizes the extinction rate of local groups¹⁶) and species selection (the idea that adaptation is best explained in terms of differential extinction of species¹⁷) have their advocates, but (with good reason) are not currently popular in most adaptive contexts.

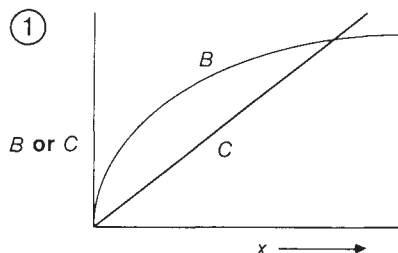
The optimization criterion is often an indirect measure of fitness. For example, when analysing animal gaits, Maynard Smith and Savage¹⁸ assumed that energy expenditure at a given speed would be minimized (or conversely, that speed for a given energy expenditure is maximized). In the case of animals foraging for food, it is often assumed⁵ that natural selection will maximize the net rate of energy intake, with the number of offspring (fitness) increasing with total calorific intake over a long period of foraging. Sometimes it is assumed that fitness is directly proportional to the indirect criterion used: however, this will often not be so, and nonlinearity may affect the quantitative predictions of the model (see Box 1.4).

Once the optimization criterion has been chosen, assumptions have to be made about the fitness consequences (or 'payoffs'—a term borrowed from game theory^{2,19}) of the different strategies, which may involve construction of mathematical models.

Payoffs are expressed in units of the criterion to be maximized, and are thus direct or indirect measures of fitness. They often

BOX 1 Some examples of simple (frequency-independent) optimization

(1) Benefits B or costs C of adopting strategy x . For the case of the foraging lapwing, $B(x)$ is the calorific value of prey items obtained after each move of distance x , and $C(x)$ is the energetic cost of moving distance x . For reasons discussed in the text, we expect $B(x)$ to increase to its asymptote by the stage where the movement distance equals the diameter of the circle of perception (this function can readily be formulated explicitly), and we expect $C(x)$ to be linear increasing.



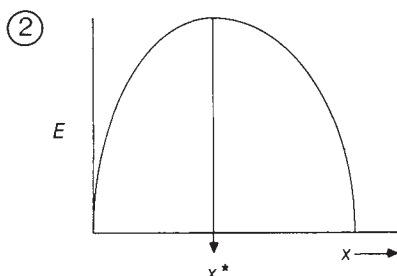
(2) The indirect fitness function. The proposed optimization criterion is the net energy gain per move, E . This is $E(x) = B(x) - C(x)$. The best movement distance, x^* , is that which maximizes E . It therefore has

$$\frac{dE(x)}{dx} = 0$$

and satisfies

$$\frac{d^2E(x)}{dx^2} < 0$$

in order that x^* is a maximum, not a minimum. This gives $B'(x^*) = C'(x^*)$; that is, the optimal movement distance occurs when the gradient of the benefit function drops to the same value as the gradient of the cost function. Biologically, the lapwing should continue to move until the marginal benefits from the next step become equal to (or less than) the marginal costs of the step.



(3) A possible way in which the indirect criterion (energy gain E) might convert into direct fitness (number of progeny W). We assume that after some large number N of movements, the total energy gained, $E = NE(x)$, is converted into total babies produced following the relationship $W(E)$. We now ask the question whether the optimal strategy under the indirect criterion will be optimal under the true criterion (fitness). The true optimum is found by maximizing $W(NE(x))$, that is, by again setting

$$\frac{\delta W(NE(x))}{\delta x} = 0$$

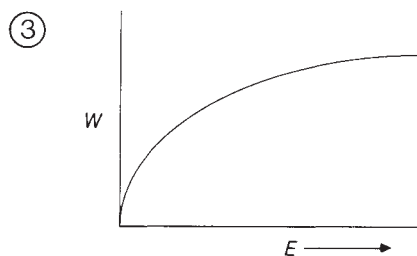
subject to

$$\frac{\delta^2 W(NE(x))}{\delta x^2} < 0.$$

This gives the result that

$$N \cdot W'(E) [B'(x) - C'(x)] = 0.$$

Because $N > 0$, then providing that W is always increasing in E (more energy means more babies), $W'(E) > 0$, so that as before $B'(x^*) = C'(x^*)$. The exact form of $W(E)$ is irrelevant, both the direct and indirect criteria give the same optimum. x^* .



(4) Charnov's marginal value theorem²⁸ relates to an animal foraging amongst a series of patches that contain prey items. The cumulative calorific gain E through foraging for time t in a patch shows diminishing returns. Time s is spent each time the animal moves between patches. The optimization criterion is the overall rate of gain from the habitat, and the question concerns the optimal time, t^* , to spend in each patch. This is found by maximizing $E(t)/(s+t)$, which gives the result that

$$E'(t^*) = E(t^*)/(s+t^*)$$

which can be demonstrated graphically by drawing the tangent to $E(t)$ as shown.

Whether t^* is the true optimum depends on whether babies are produced at the end of a long time foraging (as Charnov's model intended), or whether they are produced at the end of each foraging bout, before the animal moves on to the next patch. If the animal has a fixed (long) time T available it will visit $T/(s+t)$ patches before it must finally convert its total gain G (where $G = E(t) \cdot T/(s+t)$) into babies W ; the true optimum is therefore found by differentiating $W(G)$ with respect to t . The result is

$$T \cdot W'(G) [E'(t^*) - E(t^*)/(s+t^*)] = 0.$$

This gives the same conclusion as in the lapwing model above; that is, provided that $W'(G) > 0$ (the more total energy gained the more babies are produced), t^* is the same for the direct and indirect criteria.

But this will not be so if babies are produced after foraging in each patch. In this case the total babies produced in time T is

$$W(E(t)) \cdot T/(s+t).$$

Differentiating this fitness function with respect to t now gives the result that

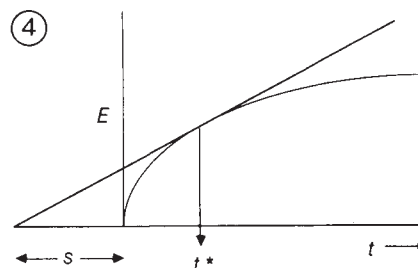
$$W'(E) \cdot E'(t^*) = W(E(t^*))/(s+t^*).$$

This is again a marginal value solution, but it stipulates that the tangent must be drawn to the direct fitness function, $W(E(t^*))$. It is easy to see that this is not equivalent to the marginal value solution for the indirect fitness function:

$$E'(t^*) = E(t^*)/(s+t^*)$$

unless the conversion between energy E and babies W is linear, that is, unless $W = aE(t)$ where a is a positive constant so that $W'(E) = a$. For this case then, the indirect criterion will result in an incorrect prediction of t^* unless it relates linearly to the direct criterion.

In general, it should be possible to investigate analytically whether—and under what conditions—an assumption of linearity will affect the predictions of a model.



vary continuously if the strategy set is continuous, and can be termed 'fitness functions'. These usually depend on trade-offs between the counteracting costs and benefits of strategy changes. For example, in his sex-ratio theory, Fisher¹⁰ assumed that all parents have equal resources to spend on the production of offspring, and that as a consequence, more sons would mean fewer daughters, and vice versa.

For specific models, fitness functions are often generated empirically, either experimentally, or by using the available natural variation, but both may pose serious problems²⁰. As an illustration of the difficulties, consider the problem of optimal clutch size in birds. Lack^{21,22} argued that the bigger the clutch, the more chicks that could be hatched. But the more chicks, the lower the chance that a given chick would survive, because the parents' input of food to the nest is limited. To build up a fitness function from natural variation, Lack recorded the initial clutch size in a nest (the variation defined the strategy set), and later counted the number of young fledgling (the fitness measure). The assumption here is that the natural variation is merely maladaptive noise. But this need not be so: parents with different resources might have different optimal clutch sizes. To overcome this problem, it is possible to alter experimentally the size of a series of clutches that were initially equal by adding or subtracting eggs as soon as egg-laying has finished, so as to measure directly the payoffs associated with different clutch-size strategies; but even this does not account for the costs to the parent of producing those eggs²⁰. In reality, it is seldom certain that payoffs and fitness functions are true measures of the fitness of the strategies studied.

At least for general models, this is not a problem; payoffs must make sense qualitatively, but need not be defined precisely. Often, all that is required is that given relationships have a certain shape (for example, survival prospects of an offspring increase monotonically with the amount of parental care it receives²³), or that relative payoffs are correctly ordered (for example, survival prospects of an offspring are higher when two parents care for it than when only one does²⁴).

Once payoffs to the strategies have been stated, the optimal solution(s) are deduced by an appropriate analytical technique. This may involve no more than differentiating the fitness function and solving for the maximum (a simple example is worked through in Box 1). But there are a great many optimization techniques, many of them are borrowed from economics and engineering (see ref. 25). An important recent advance which allows analysis of optimal short-term, state-dependent decision-making is the application of stochastic dynamic programming in animal behaviour^{7,26}.

The final step in the optimality approach is to test the predictions, quantitatively or qualitatively, against the observations. If they fit, then the model may really reflect the forces that have moulded the adaptation. If they do not, we may have misidentified the strategy set, or the optimization criterion, or the payoffs; or the phenomenon we have chosen may not in fact any longer be adaptive. By reworking our assumptions, we modify our model and revise and retest the predictions. This has been criticized as being an iterative procedure leading inevitably to a fit. But this is how science works; theories can only be discarded when they are disproven or found to be unrealistic.

Two types of optimum

Optimization may be simple (frequency-independent) or competitive (frequency-dependent) depending on whether or not the optimum for an individual is affected by what other individuals do.

For an example of simple optimization consider the foraging behaviour of lapwings, which when searching for food typically move a few paces before pausing to look for their insect prey which are eaten if seen²⁷. Over the observed range, the time spent stepping is trivial relative to pausing, but energy is expended at each step. Thus to deduce the optimal amount of move-

ment between each bout of pausing, a plausible optimization criterion is the net gain of energy per bout.

If the distance moved is small, most of the ground will already have been inspected during the previous bout, making the discovery of a new food item unlikely. When the distance moved equals the diameter of the visual field, further movement does not help. The expected energy gain will show an asymptotic relationship with distance moved x , whereas energetic costs are assumed to rise linearly with x (Box 1.1). Net energy gains are found by subtracting costs from benefits (Box 1.2). Assuming that it pays to move at all, the optimal strategy (x^*) is the distance that maximizes energy gain; it can be calculated by differentiating the fitness function as shown in Box 1. This is a simple (frequency-independent) optimum because the best strategy depends only on the energetic costs of moving each unit of the diameter of the visual field, and not on the strategies of other lapwings.

If the indirect criterion is not, as assumed, linearly related to fitness (see Box 1.3), the predictions may or may not be reliable, depending on the nature of the model. In the above example, linearity is not essential, but Box 1.4 shows that one version of the marginal value theorem^{13,28}—which predicts how long to stay in a food patch—does require the assumption of linearity, though Charnov's²⁸ original version does not.

By way of contrast, the optimal foraging rate of an individual in a flock of birds searching for berries in a tree is not simple, but depends on the behaviour of other individuals. Thus, whereas a solitary bird may consume berries at a slow rate because it has no competitors, in a flock, any bird that forages at a higher speed than the rest obtains a larger share of the berries. The best strategy now depends critically on the foraging rates of other birds, and payoffs now depend on the frequencies of strategies in the population. It is possible, under such circumstances, to seek a competitive equilibrium or evolutionarily

BOX 2 Conditions for a strategy I to be an ESS against all alternative strategies J (refs 1, 2, 31)

There are two conditions:

$$W(I, \text{Pop}_I) \geq W(J, \text{Pop}_I) \text{ for all } J \quad (1)$$

where $W(I, \text{Pop}_I)$ is the payoff to a single I strategist in a population of I strategists, and $W(J, \text{Pop}_I)$ is the payoff to a rare mutant J strategist in the same population. For continuous fitness functions where the ESS is a unique competitive optimum, we can solve for the ESS after the method outlined in Box 2.

But suppose that

$$W(I, \text{Pop}_I) = W(J, \text{Pop}_I) \quad (2a)$$

that is, some mutant strategy J can achieve a payoff equal to the payoff of I in a population of I . Then we require that

$$W(I, \text{Pop}_{I\Delta J}) > W(J, \text{Pop}_{I\Delta J}) \quad (2b)$$

in which $\text{Pop}_{I\Delta J}$ is a population of I in which there is a sufficiently small proportion of J .

Condition 2b simply ensures that if J achieves an equal payoff when rare, it is lost by selection as soon as its frequency increases sufficiently by genetic drift. It has a special relevance when the ESS is a mixed strategy (for example, play strategy, S_1 with probability p_1 , play strategy S_2 with probability p_2 , ... et cetera, where $p_1, p_2, \dots$ et cetera, are prescribed by the ESS). In this case, any one of the component pure strategies (for example, S_1 or S_2 et cetera) played with probability 1, or any rare alternative mixed strategy (in which the same pure strategies are played with different probabilities from the ESS) will obey equation 2a, because a condition for a mixed strategy to be an ESS is that each of the component strategies is equally rewarding when played in a population exhibiting the mixed ESS.

The study of ESS theory and its application to biological problems has expanded rapidly during the past decade or so, and owes much to the mathematics of game theory^{2,29}. An ESS which obeys condition 1 is equivalent to the game-theoretical concept of a Nash equilibrium, though the ESS condition 2 has no parallel in game theory¹⁹.

BOX 3 An example of competitive optimization with continuous fitness functions

(1) Fitness W achieved by mutants deviating continuously in strategy x , when the population is fixed at either x_1 , or x^* , or x_2 . In the bird flock example in the text, x is a foraging speed, and energetic costs of foraging increase with speed. But birds forage in flocks, and so individuals with higher than average foraging rates gain a higher than average proportion of the berries in each tree. It is assumed that the energetic costs of moving to another tree are high. A more explicit formulation of this type of model has been analysed by Parker³². In a population that plays strategy x_1 , a bird that deviates by playing higher rates of foraging achieves a higher payoff, W . The best strategy in the x_1 population is shown by the circled dot. In a population where birds play the high foraging rate x_2 , the reverse applies, deviations towards lower rates of foraging are favoured. Hence neither x_1 nor x_2 can be stable; selection will tend to favour values of x as indicated by the arrows. In the ESS population, x^* , any mutants deviating from the foraging rate x^* receive lower payoffs: x^* satisfies ESS condition 1. Following the notation in condition 1, we can solve for x^* by setting

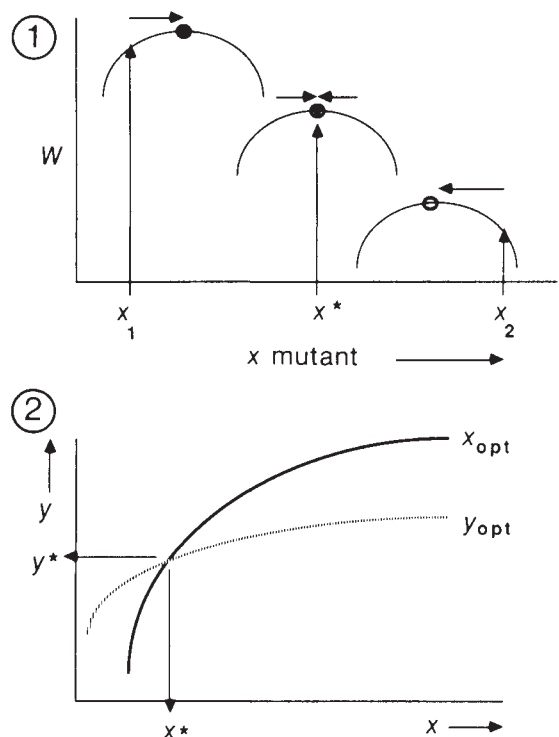
$$\left. \frac{\delta W(x, \text{Pop}_{x^*})}{\delta x} \right|_{x=x^*} = 0$$

subject to

$$\left. \frac{\delta^2 W(x, \text{Pop}_{x^*})}{\delta x^2} \right|_{x=x^*}$$

This technique guarantees only local stability; it will be necessary to check that extreme mutants cannot invade.

(2) Games in which one continuous strategy set plays against another continuous strategy set. There are several biological games in which one strategy always plays against a different strategy. Examples include parent-offspring conflict, where the offspring strategy may be a begging level and the parental strategy may be a parental investment level²³, and sexual conflict, where the male strategy may be a courtship persistence level and the female strategy a courtship resistance level. Let the two opponents be X and Y , and let their continuous strategies be x and y . The graph shows the line of x_{opt} —the best strategies for X to play against all possible strategies played by Y , and the line of y_{opt} —the best replies of Y against X . The ESS is the pair of strategies x^*, y^* ; these occur at the intersection of the two lines of best replies. In an x^*, y^* population, mutants deviating unilaterally in y or x will receive lower payoffs; the ESS will again satisfy condition 1. Following



the same notation as in (1), we can solve for x^* by setting

$$\left. \frac{\delta W(x, \text{Pop}_{x^*, y^*})}{\delta x} \right|_{x=x^*} = 0$$

subject to

$$\left. \frac{\delta^2 W(x, \text{Pop}_{x^*, y^*})}{\delta x^2} \right|_{x=x^*} = 0.$$

Strategy y^* is solved in the parallel fashion (by interchanging x and y , and evaluating at $y = y^*$).

stable strategy (ESS) (ref. 2) using mathematical techniques derived from game theory²⁹. The conditions for a strategy to be an ESS were first proposed by Maynard Smith and Price²⁹ and are outlined in Box 2.

The fitness function for ESS analysis prescribes the payoffs to rare mutants which deviate in strategy from the rest of the population. In our example of the bird flock, we anticipate that: (1) if the average foraging rate is low, a mutant that forages faster does better because it gets a bigger share of the berries; (2) if the average foraging rate is high, a mutant that forages slower does better because it spends less on foraging (see Box 3.1). In the ESS population, a mutant that forages at the average rate for the population does best; that is, the ESS is uninvadable by any alternative mutant strategy as defined in Box 2. In the example of the bird flock, each individual has equal strategic possibilities. But the same technique can be used to analyse games between different types of individual (for example, males and females; parents and offspring) who may play quite different strategies (see Box 3.2).

An ESS is not the strategy that maximizes fitness in a population sense (though this can be the case). In Box 3.1, maximum population fitness is achieved when the foraging speed is low, but such a strategy is not a competitive optimum or ESS, because it is invadable by a mutant that forages faster. In the ESS

population, fitness is maximized only in the sense that mutants not playing the ESS do worse.

Many optimality models have globally stable equilibrium solutions towards which selection is expected to converge. Quite often, however, a given equilibrium is stable only locally (against mutants showing small deviations from it) and if an extreme mutation arises, or if the population contains high frequencies of alternative strategies, selection may drive towards a different equilibrium. Multiple equilibria are common in ESS models. Likewise, in simple optimization there may be several 'adaptive peaks' so that the outcome of selection depends critically on the starting conditions.

Can natural selection optimize?

The ability of natural selection to optimize depends on gene expression and on the mechanism of genetic change in populations. Of particular importance are the rate at which selection can alter the genetic structure, the amount of additive genetic variance present at the start of selection, gene flow (for example, that arising from immigration), the rate at which conditions change, and random effects such as genetic drift.

Most optimality models assume that strategies reproduce asexually, or if the model is mendelian, that the optimal phenotype can breed true. Pleiotropy (genes affecting multiple

traits) is assumed not to operate and strategies are allowed to replicate independently of each other. Obviously, selection cannot produce an optimum if there is no way of achieving it genetically, but for some models, it is clear that selection will get as close to the optimum as the genetic mechanism will allow³³.

If selection can produce optima it is necessary to explain the natural variability found in populations. Such variation is viewed by some (the 'balancing-selectionist' school of ecological genetics) as being selectively maintained balanced polymorphisms, and by others (the 'neutralists') as being selectively neutral. Heterogeneity and changing conditions must mean that often populations are not perched at adaptive peaks. Even when conditions are constant, selection becomes progressively weaker towards the peak of a continuous fitness function; infinite time and infinite populations would be needed to achieve the peak itself.

Environmental effects will also generate variation, and this can be important in determining the optimal genotype³⁴. For example, if a normal probability distribution of phenotypes arises from a given genotype and fitness as a function of phenotype is skewed, the optimal genotype will not correspond to the peak of the fitness function—it will lie on the side where the fall is less steep, where random deviation from the genotypic mean is less heavily penalized.

Competitive optimization poses special problems. Sometimes the predicted ESS is a mixed strategy consisting of a set of pure strategies played with probabilities prescribed by the ESS, posing various difficulties beyond our scope (but see ref. 2). Some game theoretical treatments (such as games in extensive form³⁵) can become so complex that it is difficult to choose between large numbers of candidate ESSs. Some models do not generate any optimum or ESS, instead the strategy (or gene) frequencies change cyclically or chaotically without stabilizing. If these models are founded on the correct biological assumptions, then of course optimization cannot be expected. In certain instances, there can be a locally stable optimum or attractor, outside which there is cycling or bounded chaos. The outcome then depends on the starting conditions.

Difficulties

It is often difficult to decide the information constraints for the optimization; this relates to the fine-tuning of adaptation. Selection may merely generate an optimal response to the weighted mean of a wide set of conditions. If discrimination is possible, separate adaptive optima might be expected under different conditions. Yet finer tuning would allow graded responses to a continuous environmental variable (for example, a temperature cline) such that the response curve is optimal. The ultimate adaptive response could be some function of n variables, where n is a very large number. Only the most ardent pan-selectionist would suppose that there can be no limit to the complexity of the adaptive response surface. Perceptual limits alone must impose limits on ability to achieve perfection, and the information plausibly available to the animal will critically affect the predictions of a model.

The fit demanded between the predictions of a model and observation depends on the nature of the model. The simple, 'general' type of model will contain very few assumptions, and can be expected to make only qualitatively correct predictions. But if these are sufficiently striking, the model has served its purpose. Two examples will illustrate this point. Hamilton¹¹ predicted that, if the offspring of a single female mate among themselves before dispersal (this is often the case, for example, in those parasitic hymenoptera in which a female lays many eggs in a single host), the sex ratio should be strongly female-biased. This prediction is born out. Hamilton went on to ask what is the optimal sex ratio if more than one female lays eggs in a single host. He got an answer that is qualitatively correct (the bias should be less extreme), but quantitatively inexact,

because he did not allow for the fact that the second female is able to perceive that the host is already parasitized. To give a second example, we³⁶ predicted that a contest between the owner of a resource and an interloper should be settled conventionally in favour of the former, even if there is no difference in fighting ability or prospective payoff. This prediction has also been confirmed^{37,38}, but most real contests are complicated by secondary asymmetries, for example in size³⁹ or in the value of the resource⁴⁰.

In contrast, 'specific' models are intended to describe a single species in a particular environment, and therefore can give quantitatively accurate predictions. There are many examples where optimality models have generated predictions that match the observations; we just give one example as a case study. It is the work of Schmid-Hempel, Kacelnik and Houston⁴¹ on honeybee foraging (Box 4). A worker bee visits flower patches repeatedly to gather nectar; between visits the nectar is returned to the hive. Assuming that the nectar available in the flower patch does not deplete (at least over the short term), then filling the crop at each visit would maximize the net rate of energy extraction from the food sources, which is commonly used as the optimization criterion in nectarivore foraging studies⁴². But bees often leave food sources when their crops are only partially filled⁴³, which cannot maximize the net rate of energy extraction. Models that include the metabolic costs of transport of nectar (during the patch visit and back to the hive) can predict that only part of the crop should be filled. The observed crop loads fit less well with those predicted when the optimization criterion is net energy gain per unit time than when it is net energy gain per unit energy expended (Box 4). The latter criterion would be more plausible if a worker bee's survival decreases with the amount of flight; that is, if it has a fixed energy budget for flight activity. Recent evidence suggests that foraging effort can reduce survival, though the effect is so far demonstrable only at excessive work loads^{44,45}. This study shows how the optimality approach can be used to discriminate between various alternative hypotheses (in this case, alternative optimization criteria) so as to gain better insight into the nature of adaptation (in this case, partial crop loading by worker bees) under the biological constraints (the various metabolic costs of flight, and the costs of flight on body condition).

When discriminating between different models it is vital to test both their assumptions as well as their predictions; a common error is to ignore assumptions, particularly if these have not been made explicit. For quantitative models, there is a problem in assessing the significance of a fit. Because the optimality model serves as a null hypothesis, ideally it must be shown that there are no statistical grounds for rejecting it. This poses two problems. First, although it is generally easy to estimate the variance of the data, it is usually unsatisfactory to assume that the prediction has no variance. Quantitative predictions are usually calculated from measured parameters (which clearly should be independent of the observations to be tested), each of which has its own variance; hence there is the problem of estimating the combined variance of the prediction. It is often useful to perform some sort of sensitivity analysis (as in Box 4) to determine the effects that parameter variations have on the prediction. Second, most statistical techniques are designed to tell us when to reject null hypotheses, rather than when to claim that predictions and observations are likely to be the same.

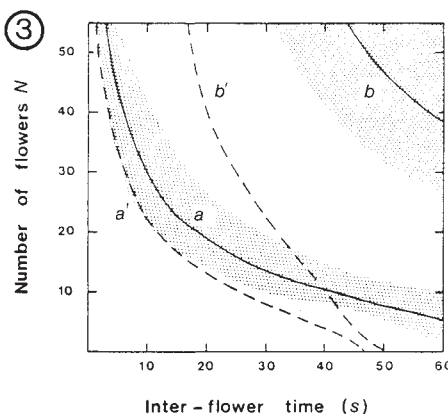
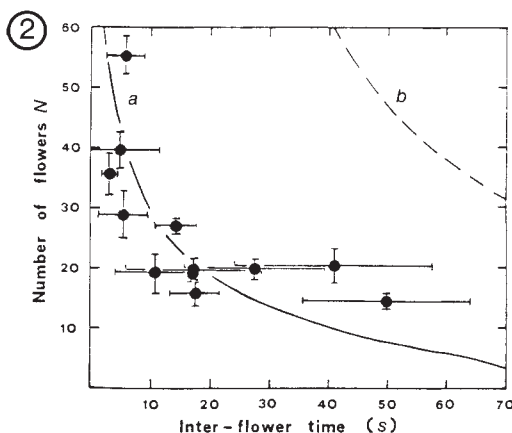
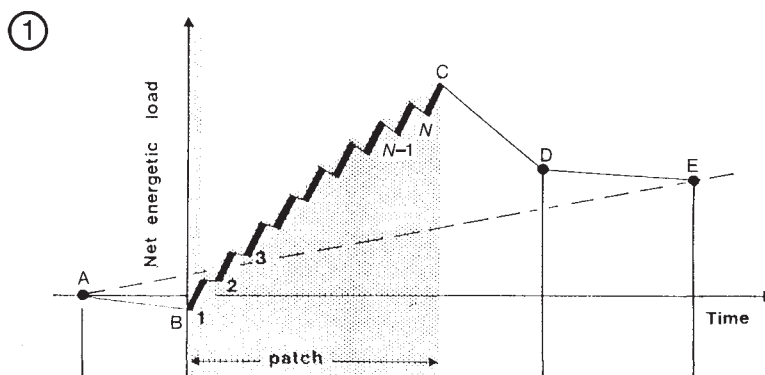
When there are several models, then the one giving the best fit to the data can be decided by some appropriate statistic. If the predictions of two models are similar; it becomes difficult to differentiate between them. If the models are themselves rather similar, and qualitative predictions alone are required, then there is probably little need to differentiate because heuristic aims will be satisfied by both. But if radically different models generate the same predictions, no progress can be made until some way is established for differentiating between them.

BOX 4 Schmid-Hempel, Kacelnik and Houston's⁴¹ study of foraging in worker honeybees

(1) The net energetic load carried by a worker bee as a function of time during the foraging cycle. The bee leaves the hive at A and encounters the first flower in the flower patch at B, having lost some energy due to the flight. It then visits a series of flowers; 1, 2, 3, ..., N , and at each gathers a fixed amount of energy in the form of nectar (shaded zone). As the crop load increases, the drop in energy due to the flight between flowers increases. The bee leaves the patch at C, after gathering from N flowers. It returns to the hive at D, having lost considerable energy due to the increased metabolic costs of flight as load increases. Between D and E the worker remains in the hive. The authors were able to calculate the gross energetic gain (G), the total energetic expenditure or loss (L) and the total time (T) per foraging cycle, each as a function of the number of flower visits (N). The slope of the broken line AE gives the net rate at which energy is delivered to the hive; if the optimization criterion is net energy gain per unit time = $(G - L)/T$, the slope of AE is maximized. For the alternative model where the optimization criterion is energy gain per unit energy expended = $(G - L)/L$, the time axis would be replaced by an axis of energy expenditure. Under each model, an optimal number of flower visits (N^*)—and hence an optimal crop loading—can be predicted. This depends on the typical inter-flower time.

(2) The number of flowers visited (N) as a function of the inter-flower time. The predicted number of flowers that should be visited under the two models is shown by the two curves in the graph; broken curve (b) is for maximization of net energy gain per time, and solid curve (a) is for maximization of net energy gain per energy expended. In both models, N^* decreases with inter-flower time, and both models predict partial crop loads at high inter-flower times (a full crop is approximately 55 flower visits). The observations for 12 different individuals are shown (●, mean; bars, standard deviation). They clearly fit curve a better than curve b , and analysis showed that the observations differ significantly from b , but not from a .

(3) Sensitivity of the model predictions to changes in the parameter values. Curves a and b are the same as in 2 above. The shaded areas around these curves show the range of predictions obtained by halving or doubling in turn each of the four parameter values for the metabolic rate for the bees (unloaded rate; increment in rate due unit load increase, flower handling rate, and hive rate). In particular, the energy gain per energy expended model appears to be very robust, and insensitive to these changes in parameter values. The broken lines (a' , b') are the predictions if the hive activity part of the foraging cycle (D to E in 1 above) is omitted from the models. This affects the quantitative conclusions but not the qualitative conclusion that the better fit is gained by the energy gain per energy expended model. The authors conclude that there are no *a priori* reasons for such an omission. Further support for this interpretation of honeybee foraging strategy has been obtained from recent physiological studies on metabolic rates⁴⁶, and the fact that altering the



volume of sugar solution available per visit to artificial flowers gives changes in number of flower visits that again fit only the energy gain per energy expended model⁴⁷.

Alternative approaches

Two other methods have been used in analysing the role of selection in evolution: the comparative method, and quantitative genetics. The former goes back to Darwin. In its modern form⁴⁸, it consists of a statistical analysis of the ecological and phenotypic characteristics of a set of species. It can be carried out 'blind', in the hope that the pattern of observed correlations will suggest a selective explanation. Alternatively, it can be used to test some specific hypothesis. For example, Clutton-Brock *et al.*⁴⁹ used the comparative method to decide between two explanations for sexual dimorphism in size in primates: that it is a consequence of competition between males for females, or that it facilitates an ecological division of labour between the sexes.

Because dimorphism is least in those species in which the breeding group is a monogamous pair, they conclude that male-male competition is the more likely explanation. The method has been extensively used to analyse behavioural, morphological and life-history evolution. It is a complement to rather than an alternative to optimization.

The approach from quantitative genetics^{50,51} depends on the measurement of genetic variances and covariances. The response of a population to selection depends on the genetic variance, and if there is a genetic covariance between traits A and B, selection on A will change B. A criticism of this approach is that genetic covariances are hard to measure, and are unlikely to be a constant and stable feature of a species. But in some

cases such covariances may reflect a fundamental physiological constraint. For example, Rose and Charlesworth⁵² found a negative covariance between longevity and female fecundity in *Drosophila*. One reason for thinking that this may be a rather stable characteristic is that there is physiological evidence⁵³ that treatments (including 5,000 rads of X rays) that reduce egg-laying in *Drosophila* also prolong life. This example shows how the discovery of a genetic covariance can indicate the presence of a developmental constraint that must be taken into account in formulating an optimization model—in this case a model of life-history evolution. Most often, however, physiological experiments and interspecies comparisons may be easier ways of discovering such constraints.

To summarize on alternative approaches, quantitative genetics can help to identify the constraints needed for a satisfactory optimization model, and the comparative method is a powerful tool for testing the predictions of such models.

Future prospects

It is not surprising that optimality theory in evolution has antagonists as well as protagonists. Some of the critics of optimality theory—often theoretical geneticists—have tended to focus on the genetic problems of reaching optima. On the other hand, adaptationists such as behavioural ecologists have tended to proceed with faith in the power of selection, using optimization models as a tool to understand adaptation. We see both disciplines as perfectly worthy of study, and not contradictory. A start has been made in comparing the predictions of game

theoretic and diploid genetic models³³, but further work is needed. A problem of particular interest from this point of view is that of the mate choice aspect of sexual selection. This would seem to be an ideal topic for analysis using game theory, because the optimal strategy for each sex depends on the behaviour of the other. So far, most analyses have involved explicit genetic models^{54–56}, but recently Grafen⁵⁷ has given a game theoretical analysis.

Similarly, in recent years students of animal behaviour have tended to belong to one of two schools. One group have sought the 'function' of behaviour and the selective forces responsible for its evolution. Others have had a primarily 'causal' or mechanistic interest, seeking to understand how the brain, together with sensory and motor organs, generates behaviour. Users of optimization theory belong to the first of these two schools, but they are forced to make assumptions about constraints: that is, assumptions about what is physiologically possible. Again, further progress is therefore likely to depend on collaboration rather than antagonism between the two schools.

In conclusion, optimization is one of the most powerful and elegant tools for understanding adaptation. None of its inherent problems call for its abandonment⁵⁸, though this has been advocated⁵⁹. But we should be aware of the difficulties. □

G. A. Parker is in the Department of Environmental and Evolutionary Biology, University of Liverpool, L69 3BX, UK. J. Maynard Smith is in the School of Biological Sciences, University of Sussex, Falmer, Brighton, Sussex, BN1 9QG, UK.

- Maynard Smith, J. A. *Rev. Ecol. Syst.* **9**, 31–56 (1978).
- Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge University Press, 1982).
- Alexander, R. M. *Optima for Animals* (Arnold, London, 1982).
- Krebs, J. R. & Houston, A. I. in *Ecological Concepts: the Contribution of Ecology to an Understanding of the Natural World* (ed. Cherratt, J. M.) 309–338 (Blackwell, Oxford, 1989).
- Stephens, D. W. & Krebs, J. R. *Foraging Theory* (Princeton University Press, Princeton, 1986).
- Schmid-Hempel, P. in *Population Biology* (eds Wöhrmann, K. & Jain, S.) 321–347 (Springer, Berlin, 1989).
- Mangel, M. & Clark, C. W. *Dynamic Modelling in Behavioural Ecology* (Princeton University Press, Princeton, 1988).
- Gould, S. J. & Lewontin, R. C. *Proc. R. Soc. B* **205**, 581–598 (1979).
- Nisbet, R. M. & Gurney, W. S. C. *Modelling Fluctuating Populations* (Wiley, Chichester, 1982).
- Fisher, R. A. *The Genetical Theory of Natural Selection* (Oxford University Press, London, 1930).
- Hamilton, W. D. *Science* **156**, 477–488 (1967).
- Charnov, E. L. *The Theory of Sex Allocation* (Princeton University Press, Princeton, 1982).
- Parker, G. A. & Stuart, R. A. *Am. Nat.* **110**, 1055–1076 (1976).
- Caswell, H. in *Ecological Concepts* (ed. Cherratt, J. M.) (Blackwell, Oxford, 1989).
- Hamilton, W. D. *J. theor. Biol.* **7**, 1–52 (1964).
- Wynne-Edwards, V. C. *Evolution through Group Selection* (Blackwell, Oxford, 1986).
- Vrba, E. S. *Syst. Zool.* **33**, 318–328 (1984).
- Maynard Smith, J. & Savage, R. J. G. *Zool. J. Linn. Soc.* **42**, 603–622 (1956).
- Parker, G. A. & Hammerstein, P. in *Evolution: Essays in Honour of John Maynard Smith* (eds Greenwood, P. J., Harvey, P. H. & Slatkin, M.) 73–94 (Cambridge University Press, Cambridge, 1985).
- Atkinson, D. in *Behavioural Ecology: Ecological Consequences of Adaptive Behaviour* (eds Sibly, R. M. & Smith, R. H.) 99–104 (Blackwell, Oxford, 1985).
- Lack, D. *Ibis* **89**, 309–352; **90**, 25–45 (1947).
- Lack, D. *The Natural Regulation of Animal Numbers* (Clarendon, Oxford, 1954).
- Parker, G. A. & Macnair, M. R. *Anim. Behav.* **27**, 1210–1235 (1979).
- Maynard Smith, J. *Anim. Behav.* **25**, 1–9 (1977).
- Clark, C. W. *Mathematical Bioeconomics* (Wiley, New York, 1976).
- Houston, A., Clark, C., McNamara, J. & Mangel, M. *Nature* **332**, 29–34 (1988).
- Pearson, R. G. & Parker, G. A. *J. nat. Hist.* **7**, 573–589 (1973).
- Charnov, E. L. *Theor. Populat. Biol.* **9**, 129–136 (1976).
- Von Neumann, J. & Morgenstern, O. *Theory of Games and Economic Behaviour* (Princeton University Press, Princeton, 1953).
- Maynard Smith, J. & Price, G. R. *Nature* **246**, 15–18 (1973).
- Parker, G. A. in *Behavioural Ecology: an Evolutionary Approach* 2nd ed. (eds Krebs, J. R. & Davies, N. B.) 30–61 (Blackwell, Oxford, 1984).
- Parker, G. A. in *Behavioural Ecology: Ecological Consequences of Adaptive Behaviour* (eds Sibly, R. M. & Smith, R. H.) 33–58 (Blackwell, Oxford, 1985).
- Maynard Smith, J. *Am. Nat.* **117**, 1015–1018 (1981).
- Yoshimura, Y. & Shields, W. M. *Evol. Ecol.* **1**, 125–138 (1987).
- Selten, R. *Math. Soc. Sci.* **5**, 269–363 (1983).
- Maynard Smith, J. & Parker, G. A. *Anim. Behav.* **24**, 159–175 (1976).
- Davies, N. B. *Anim. Behav.* **26**, 138–147 (1978).
- Kummer, H., Götz, W. & Angst, W. *Behaviour* **49**, 62–87 (1974).
- Sigurjónsdóttir, H. & Parker, G. A. *Behav. Ecol. Sociobiol.* **8**, 219–230 (1981).
- Reichert, S. E. *Behav. Ecol. Sociobiol.* **6**, 121–128 (1979).
- Schmid-Hempel, P., Kacelnik, A. & Houston, A. I. *Behav. Ecol. Sociobiol.* **17**, 61–66 (1985).
- Pyke, G. H. *Oecologia* **36**, 281–293 (1978).
- Nunez, J. *J. apic. Res.* **21**, 139–150 (1982).
- Schmid-Hempel, P. & Wolf, T. *J. Anim. Ecol.* **57**, 500–521 (1988).
- Wolf, T. H. & Schmid-Hempel, P. *J. Anim. Ecol.* **58**, 943–954 (1989).
- Wolf, T. H., Schmid-Hempel, P., Ellington, C. P. & Stevenson, R. D. *Funct. Ecol.* **3**, 417–424 (1989).
- Schmid-Hempel, P. *J. Anim. Ecol.* **56**, 209–218 (1987).
- Pagel, M. D. & Harvey, P. H. *Q. Rev. Biol.* **63**, 413–416 (1988).
- Clutton Brock, T. H., Harvey, P. H. & Rudder, B. *Nature* **269**, 797–800 (1977).
- Lande, R. *Evolution* **34**, 402–416 (1979).
- Lande, R. & Arnold, S. J. *Evolution* **37**, 1210–1226 (1983).
- Rose, M. & Charlesworth, B. *Nature* **287**, 141–142 (1980).
- Lamb, M. J. *J. Insect Physiol.* **10**, 487–497 (1964).
- Lande, R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3721–3725 (1981).
- Kirkpatrick, M. *Evolution* **36**, 1–12 (1982).
- O'Donald, P. *Genetic Models of Sexual Selection* (Cambridge University Press, Cambridge, 1980).
- Grafen, A. *J. theor. Biol.* **144**, 473–546 (1990).
- Stearns, S. & Schmid-Hempel, P. *Oikos* **49**, 118–125 (1987).
- Pierce, G. J. & Ollason, J. G. *Oikos* **49**, 111–118 (1987).

ACKNOWLEDGEMENTS. We are especially grateful to Steve Stearns, Alasdair Houston, Paul Schmid-Hempel, David Harper and Leigh Simmons for kindly commenting on various drafts of the manuscript.

Evidence for early Proterozoic plate tectonics from seismic reflection profiles in the Baltic shield

BABEL Working Group*

Plate tectonics provides the linking framework for all tectonic and magmatic activity seen today, but it is not known when plate tectonics first developed on Earth. New deep seismic reflection and co-incident refraction profiles across an exposed, 1.89-Gyr-old volcanic arc complex show a 10-km-thick offset in the Moho and bivergent reflectors in the crust, which were most probably created by plate convergence, subduction and accretion during the Early Proterozoic. Hence, plate tectonic models seem to be applicable for at least the second half of Earth's history.

ONE of the most fundamental uncertainties in Precambrian geology is the type and timing of secular changes in tectonic style during Earth's history. Although plate tectonics is almost ubiquitously used to interpret both active tectonics and the Phanerozoic rock record, from what time onward is it appropriate to use plate tectonic concepts to help understand the Precambrian geological record? Considerable effort has gone into searching for the oldest ophiolites as evidence for ocean crust formed at mid-ocean ridges, and for the oldest blueschists as evidence for the high-pressure/low-temperature metamorphism characteristic of subduction-zone environments. The oldest complete ophiolite reported is the 2.0-Gyr Purtuniqu ophiolite, Quebec¹. The oldest blueschists known, from western China, are 0.7 Gyr old².

To the above geological evidence for Proterozoic subduction-zone tectonism, we now add evidence from seismic reflection profiles for collisional plate tectonics in the 1.93- to 1.86-Gyr orogenic domains in northern Sweden and Finland. Our combination of seismic reflection and refraction profiling defines a 10-km-deep crustal root zone at the Moho, beneath which bright reflectors extend to about 70 km depth in the mantle, and above which intra-crustal reflectors dip to the north and converge in the lower crust with south-dipping reflectors, giving a bivergent pattern. The geometry of the root zone at the Moho and of the bivergent crustal reflectors suggests that deep crustal underthrusting from the south drove northeast-directed thrusting and

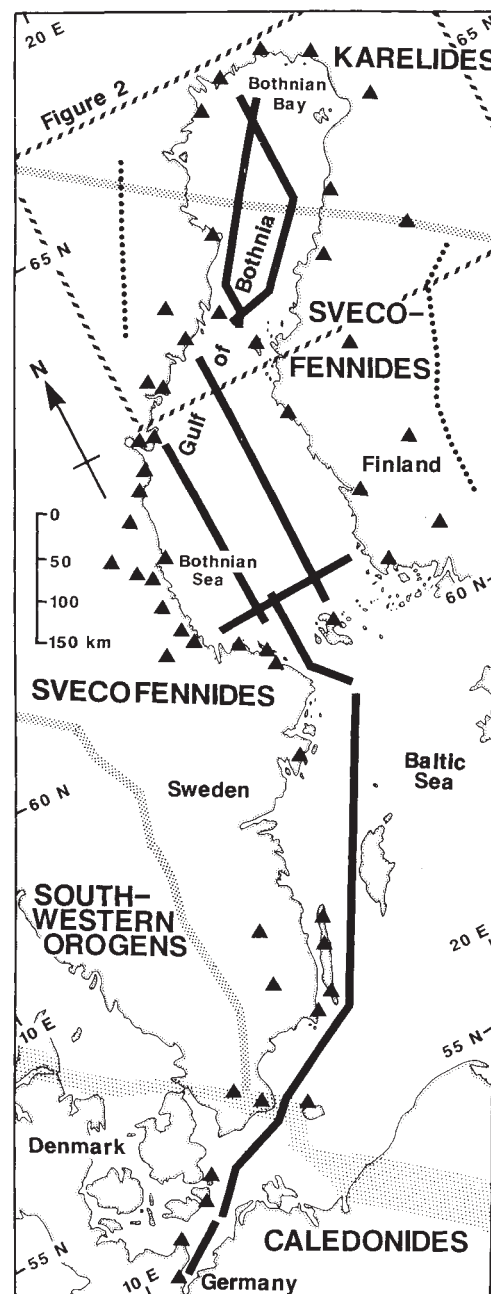


FIG. 1 Location map of reflection profiles (thick solid lines) recorded during Project BABEL and seismic stations (triangles) used to record the marine seismic source at wide angles. Stippled lines, major tectonic boundaries. Dotted lines, location of magnetotelluric stations which located the Skellefte conductivity anomaly in Sweden³⁴ and the southern Finland conductivity anomaly in Finland^{35,36}. Hatched line bounds area of Fig. 2.

* Members of the BABEL (for Baltic and Bothnian Echoes from the Lithosphere) Working Group are: T. Dahl-Jensen, R. W. Hobbs, S. L. Klemperer, D. H. Matthews & D. B. Snyder, British Institutions' Reflection Profiling Syndicate (BIRPS), Bullard Laboratories, Cambridge University, Madingley Rise, Madingley Road, Cambridge CB3 0EZ, UK; R. Long, T. Matthews, Department of Geological Sciences, Durham University, South Road, Durham DH1 3LE, UK; D. J. Blundell, Department of Geology, RHB New College, Egham TW20 0EX, UK; C.-E. Lund, H. Palm, L. B. Pedersen & R. G. Roberts, Department of Solid Earth Geophysics, University of Uppsala, Box 556, S-75122 Uppsala, Sweden; S.-Å. Elming at the Department of Geophysics, Luleå Technical University, S-95187 Luleå, Sweden; P. Heikkinen, H. Korhonen & U. Luosto, Institute of Seismology, Helsinki University, Et Hesperiankatu 4, SF-00100 Helsinki, Finland; S.-E. Hjelt & K. Komminaho, Department of Geophysics, Oulu University, SF-90570 Oulu, Finland; J. Yliniemi at the Geophysical Observatory, Oulu University, SF-90570 Oulu, Finland; R. Meissner, P. Sadowiak & Th. Wever, Institute for Geophysics, Olshausenstrasse 40, D-2300 Kiel 1, Germany; T. Dickmann & E. R. Flueh GEOMAR Institute, Wischhofstrasse 1-3, D-2300 Kiel 14, Germany; A. Berthelsen & H. Thybo, Institut for Almen Geologi, Øster Voldgade 10, DK-1350 Copenhagen K, Denmark; N. Balling & E. Nørmark, Department of Earth Sciences, Aarhus University, Finlandsgade 6, DK-8200 Aarhus N, Denmark.

† Present address: Greenland Geological Survey, Øster Voldgade 10, DK-1350 Copenhagen K, Denmark.
‡ To whom correspondence should be addressed at the Department of Geophysics, Mitchell 366, Stanford University, Stanford, California 94305-2215, USA.

folding at shallower levels. Formation of these reflector geometries was probably related to the deformation and metamorphism of an overlying arc complex between 1.89 and 1.86 Gyr BP (before present). These data provide the clearest demonstration yet of Proterozoic plate tectonics by deep seismic methods.

Project BABEL

Project BABEL (Baltic and Bothnian Echoes from the Lithosphere) is a collaborative effort of British, Danish, Finnish, German and Swedish scientists to acquire deep-crustal reflection profiles and collinear wide-angle/refraction profiles in the Baltic Sea and the Gulf of Bothnia (Fig. 1) for the study of the structure of the Precambrian Baltic shield and its transition to the Phanerozoic crust of western Europe across the Tornquist Zone. Here we discuss only the structures associated with the early Proterozoic Svecofennian orogeny in Bothnian Bay (Fig. 2). Reflection profiles from other regions, and detailed crustal velocity and thickness profiles from the wide-angle data, will be published elsewhere.

The BABEL near-vertical data were collected and are being processed as standard BIRPS deep-crustal profiles^{3,4}. In September and October 1989 we collected 2,268 km of profiles (Fig. 1) with record lengths of 18–25 s (two-way travel time, used throughout this paper) and a stacking fold of 20–30 per 25 m common-midpoint trace, using a 120.6-l (7,359 in³), 140-bar (2,030 p.s.i.) source of 48 guns towed at 7.5 m depth and a 3-km, 60-channel streamer towed at 15 m. The preliminary stack sections (such as in Fig. 3a) were processed at sea on a MicroMAX processing system, and were then migrated using a constant 6.5 km s⁻¹ velocity (on Bullard Laboratories' Convex C-1 computer using GECO Seismic Kernel System software).

We also recorded all the shots (>30,000) from the marine airgun source at more than 50 seismic stations (ranging from single three-component seismometers to 1-km-long arrays) on islands and on land around the Gulf of Bothnia and the Baltic Sea (Fig. 1). We observed arrivals up to 700 km away, the maximum offset at which we attempted to record data. Continuous recording of the marine seismic source provided very densely sampled common-receiver gathers: Fig. 4 shows data with a 75-m source spacing recorded using the airgun shots fired for the reflection profile shown in Fig. 3a. Preliminary interpretations of the wide-angle profiles (such as Fig. 4) allow us to place the refraction Moho, normally taken to define the base of the Earth's crust, at depths or travel times roughly corresponding to the reflection Moho on the collinear reflection profile (Fig. 3b).

Geological evolution of the Baltic shield

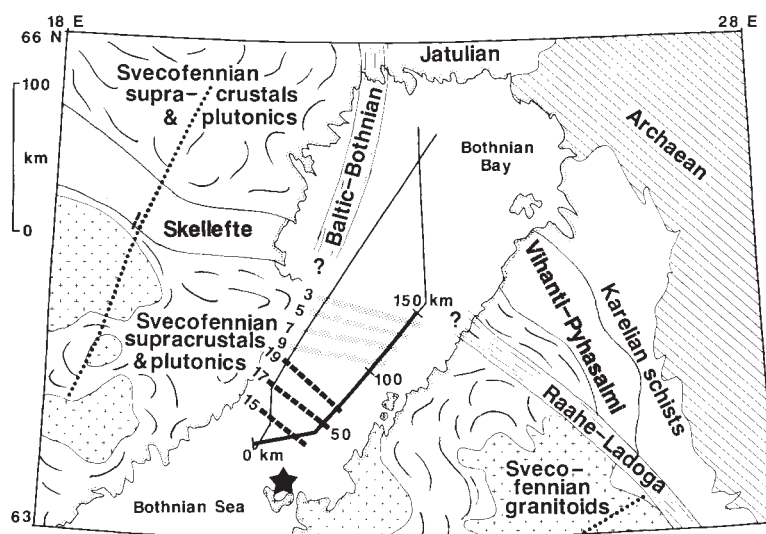
The Baltic shield evolved by the development of Proterozoic orogenic belts adjoining an Archaean nucleus^{5,6}. Archaean rocks are only exposed northeast of Bothnian Bay⁷ (Fig. 2) but, based on isotope evidence, at least relics of reworked Archaean crust must extend beneath Proterozoic metasedimentary cover and granitoid intrusions towards the Skellefte district in Sweden^{8,9} and to the Raahe-Ladoga zone in Finland¹⁰ (Fig. 2). The Archaean craton was intruded by northwest-trending dykes ~2.2 Gyr ago, which were probably associated with the formation of a passive continental margin on which the Jatulian platform carbonates (~2.1 Gyr old) and the Kalevian greywackes and turbidites (~2.0 Gyr old) were deposited⁷. North-east-vergent nappes, containing 1.96-Gyr ophiolitic fragments¹¹ within the Karelian schists, were emplaced on this margin at ~1.9 Gyr (ref. 12).

Further south, the Vihanti-Pyhäsalmi district in Finland and the Skellefte district in Sweden are 50-km-wide belts of metasediments and metavolcanics containing numerous Cu-Pb-Zn mineralizations which are believed to represent Proterozoic volcanic arcs^{13,14}. These ~1.89-Gyr (ref. 15) arcs are widely presumed to have developed above a subduction zone dipping northeast beneath the margin of the Archaean craton^{7,16,17}. The above plate tectonic view of the Svecofennian orogeny has not been universally accepted owing to the lack of unequivocal observations supporting plate tectonics (such as any complete ophiolite), and non-plate-tectonic models have also been advanced^{18,19}.

During metamorphism and deformation of the Skellefte arc at 1.88–1.86 Gyr (ref. 15), including folding about northwest-trending axes with vertical to south-dipping axial planes¹³, continued northeast-directed subduction beneath the northeastern continental area is thought to have been the cause of the widespread intrusion of Svecofennian granitoids and extrusion of volcanics onto the Archaean-Proterozoic continent between 1.89 and 1.86 Gyr^{8,14}. The Skellefte and Vihanti-Pyhäsalmi rocks, and these later volcanics and overlying sedimentary rocks, form part of the widespread Svecofennian supracrustal sequences (Fig. 2). The present exposure of rocks in the greenschist and low-pressure-amphibolite facies in the Skellefte district and of rocks in the amphibolite to low-pressure-granulite facies in the Vihanti-Pyhäsalmi district¹⁴, implies several kilometres of uplift and erosion before deposition of the youngest Svecofennian supracrustals at 1.8 Gyr (ref. 20) which are only weakly metamorphosed²¹.

The Svecofennides, extending 1,000 km to the south of the Skellefte and Vihanti-Pyhäsalmi districts, consist of belts of metavolcanics and metasediments intruded by calc-alkalic

FIG. 2 Geological summary map for Bothnian Bay^{12,17,25,45,54}. Solid lines in Bothnian Bay show the BABEL reflection profiles. Emboldened segment of profile numbered 0–150 km shows where data in Fig. 3 were collected. Star shows position of receiver for data in Fig. 4. Thick dashed lines at sea are contours of travel time (15–19 s) to deepest mantle reflectors after migration. Stippled lines are contours of travel time (3–9 s) to shallowest south-dipping reflectors after migration. Width of dashed and stippled lines reflects uncertainty in correlations between the reflection profiles. Dotted lines on land show the locations of the magnetotelluric stations. Short dashed line at the southern border of the Skellefte zone shows the location of the 10-km-long seismic profile⁴⁴.



plutonic rocks⁶. This crust largely formed between 1.9 and 1.8 Gyr BP²² (recent dating of zircon cores, however, shows the existence of a Proterozoic crustal component, as old as 2.1 Gyr (ref. 23), and intrusion of late Svecofennian granites continued as late as 1.78 Gyr BP²²).

Later in the Proterozoic, the Archaean-Proterozoic boundary zone was cut by the near-vertical Raahe-Ladoga zone²⁴ and the Baltic-Bothnian megashear²⁵. The Skellefte and Vihanti-Pyhäsalmi districts are probably along-strike equivalents¹⁷, although there may have been later displacement between them along the cross-cutting strike-slip fault systems.

Since 1.77 Gyr BP the crust beneath Bothnian Bay has been rather stable. Up to 1 km (maximum preserved thickness) of ~1.3-Gyr Jotnian sandstones were deposited over the central Baltic shield, including in and around Bothnian Bay^{26,27}. South of Bothnian Bay, dolerites of regional extent and cumulate thickness of a few hundred metres were intruded at ~1.25 Gyr (ref. 28). During the early Palaeozoic at least 375 m (maximum preserved thickness in the Gulf of Bothnia) of Cambro-Ordovician shelf sediments were deposited^{26,27}.

Previous geophysical investigations

Refraction experiments have been recorded on land in Sweden and Finland but not in the Gulf of Bothnia, so existing Moho maps are attempts to interpolate across a 200-km-wide gap in the data. Profiles across the Raahe-Ladoga zone in Finland show a Moho depth in excess of 55 km, becoming gradually shallower to the north and south^{29,30}. A north-south refraction profile (FENNOLOLA: Fennoscandian long-range profile) across the Skellefte district, ~100 km west of Bothnian Bay, shows either a 5-km step in the Moho³¹, or a gradual shallowing

of the Moho from ~50 km depth west of the Bothnian Sea to ~42 km depth northwest of Bothnian Bay^{30,32}.

Both NE-SW-trending and NW-SE-trending residual gravity anomalies are observed over most of the Baltic shield, and a causal connection has been suggested between the NW-SE trend and density structures generated by Precambrian subduction and collision³³.

Magnetotelluric measurements on land in Sweden, made ~100 km west of Bothnian Bay (Figs 1 and 2), detected a conductivity anomaly south of the Skellefte district that dips 20° to the northeast to 30 km depth, strikes 145° towards Bothnian Bay, and has a magnitude of ~4,000 S (ref. 34; the Skellefte anomaly). The Oulu conductivity anomaly in the area of the Vihanti-Pyhäsalmi zone has a similarly large conductance, up to 20,000 S^{35,36}, but trends north-south and is not known to extend beneath Bothnian Bay. It has been suggested³⁶ that the Skellefte anomaly may continue directly across Bothnian Bay to connect with the southern Finland conductivity anomaly which also dips to the northeast but which has a smaller conductance, ~500 S, where it has been detected ~200 km southeast of Bothnian Bay^{35,36} (Figs 1 and 2).

New seismic data

Figure 3 illustrates the southern part of BABEL profile 4 (migrated), the eastern of two northeast-trending profiles recorded 0–50 km apart in Bothnian Bay (Fig. 2). The following features, visible in Fig. 3a and keyed by letters to the line drawing (Fig. 3b), can all be correlated between these two profiles and so can be shown to have trends varying from WNW-ESE to NW-SE (Fig. 2).

(1) At the southern end of the profile the prominent crustal

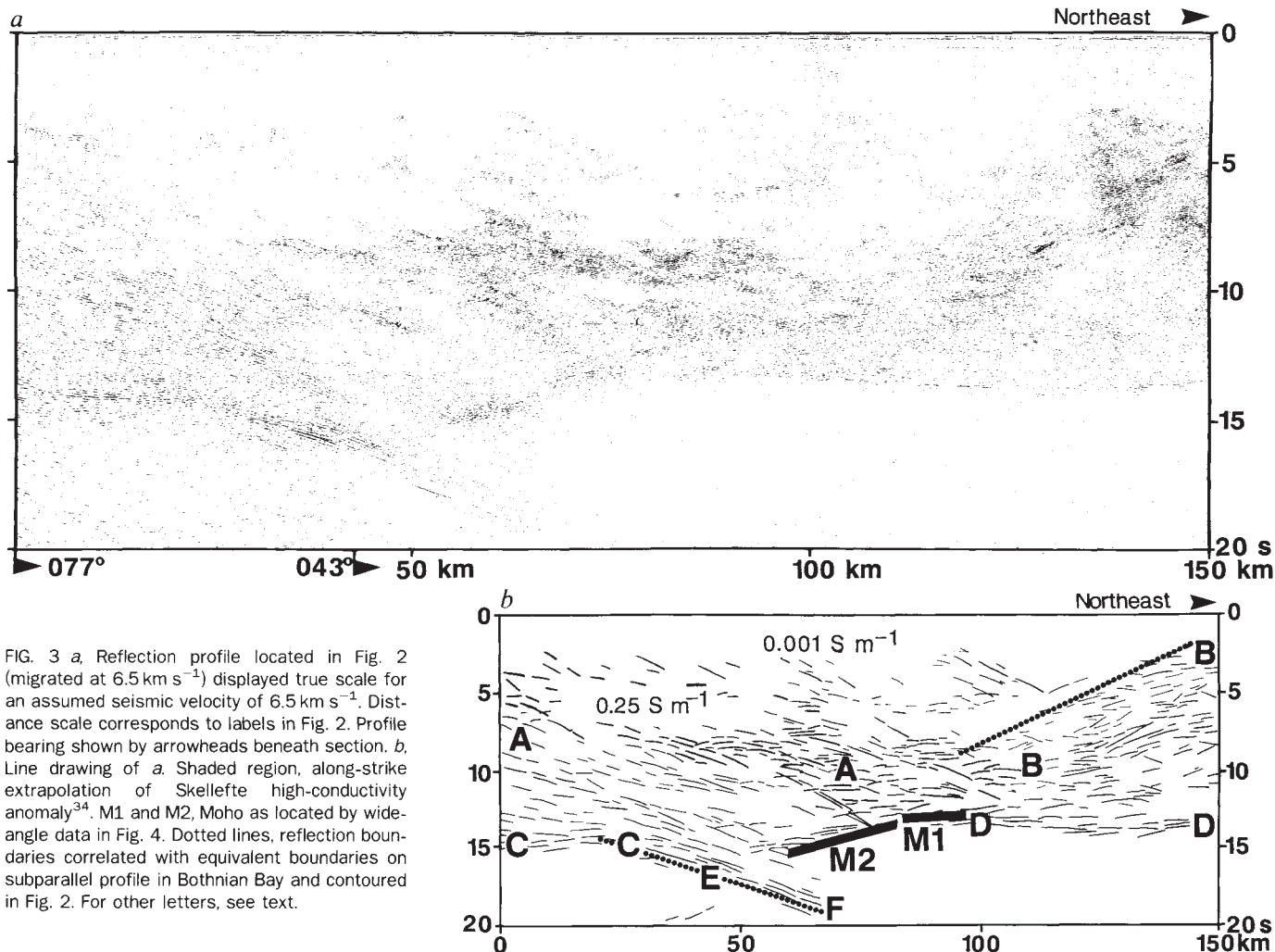


FIG. 3 a, Reflection profile located in Fig. 2 (migrated at 6.5 km s⁻¹) displayed true scale for an assumed seismic velocity of 6.5 km s⁻¹. Distance scale corresponds to labels in Fig. 2. Profile bearing shown by arrowheads beneath section. b, Line drawing of a. Shaded region, along-strike extrapolation of Skellefte high-conductivity anomaly³⁴. M1 and M2, Moho as located by wide-angle data in Fig. 4. Dotted lines, reflection boundaries correlated with equivalent boundaries on subparallel profile in Bothnian Bay and contoured in Fig. 2. For other letters, see text.

reflectors dip northeast (A on Fig. 3b), whereas in the northern part of the profile the reflective bands in the crust mostly dip south-southwest (B). Structure contours, in travel time to the top of the SSW-dipping reflectors (dotted line above B in Fig. 3b), are shown in Fig. 2.

(2) The reflection Moho, defined as the deepest high-amplitude reflection at travel times (depths) approximately commensurate with other estimates of crustal thickness³⁷, is at ~15 s on the southern end of the profile (C) and at ~13.5–14 s in the northern part of the profile (D). The wide-angle data (Fig. 4) show reflections (M1 and M2) from Moho segments marked M1 and M2 on Fig. 3b, confirming that the reflection Moho beneath Bothnian Bay is essentially at the depth defined by wide-angle reflection and refraction measurements, as seems to be true on a world-wide basis³⁸. Equivalent tests will be possible using the data from about 20 other wide-angle recorders positioned around the northern Gulf of Bothnia (Fig. 1) to record these near-vertical reflection profiles.

(3) In the southern part of the profile the reflection Moho dips northeast from 15 s (C) to 17 s (E) then rises up to 13.5 s (D). Beneath this triangular zone other reflectors (F) dip 20° to 25° northeast to nearly 20 s, or ~70 km depth. Structure contours in travel time to the base of these northeast-dipping reflectors (dotted line C–E–F in Fig. 3b) are shown in Fig. 2.

Thus, the reflection data show a clear downward convergence of two distinct bands of crustal reflectors, dipping 20° to 30° northeast in the south of Bothnian Bay but dipping 20° to 30° southwest further north. The upper part of the northeast-dipping reflectors correlates well with the along-strike projection, from the northwest, of the Skellefte conductivity anomaly³⁴. The northeast-dipping reflectors in the crust (A) and in the mantle (F) bracket a zone of pronounced Moho topography. Across a 50-km-wide zone (C–D) the reflection Moho shows a net change of 1–1.5 s travel time, or 3–5 km in depth if mean crustal velocities are uniform across this zone. This change in Moho depth may well correspond to the 5-km Moho step interpreted³¹ from the onshore FENNOLOGRA refraction profile in Sweden. Over the shorter distance of 25 km from E to D, there is a step of 3.5 s in travel time, or 10–15 km.

Comparison with Phanerozoic orogens

The clear Moho offset and the bivergent crustal reflectors beneath the southern Bothnian Bay form a reflection pattern that seems to be characteristic of many Phanerozoic orogens, including the Alps³⁹ and the Pyrenees⁴⁰, both of Cenozoic age, and the Appalachians⁴¹ and Caledonides^{42,43} of Palaeozoic age. In each of these orogens, all better understood than the Svecofennides, the bivergent reflections in the crust seem to mark tectonically imbricated crust on opposite sides of a suture

zone. The offset Moho and/or crustal root zone in each of these examples has been interpreted as the locus of either subduction of oceanic crust or under-thrusting of continental crust in the collision zone. The reflectivity pattern of the BABEL seismic data (Fig. 3) therefore encourages us to seek a geological interpretation (Fig. 5) consistent with our understanding of Phanerozoic orogenies.

Geological interpretation

The shallowest SSW-dipping reflectors at the north end of Fig. 3 (dotted line in Fig. 3b) project to the surface near the southern margin of the Skellefte district on the Swedish side of Bothnian Bay, and in the Raahe–Ladoga zone on the Finnish side. SSW-dipping reflectors in the upper half of the crust continue for ~50 km north of the data shown in Fig. 3. The Raahe–Ladoga zone is interpreted as a series of vertical shear zones²⁵ so we prefer to interpret the south-dipping reflectors as being related to the Skellefte district, which contains folds with south-dipping axial planes¹³. This interpretation is supported by the observation of south-dipping reflections projecting to the surface in the Skellefte district on a 10-km-long reflection profile recorded at the southern border of the Skellefte zone, ~85 km inland⁴⁴ (Fig. 2). This correlation of reflectors observed close to the Finnish coast with the Skellefte district in Sweden implies that sequential sinistral and dextral movements on the Baltic–Bothnian megashear^{25,45} produced a net offset of less than a few tens of kilometres between the BABEL reflection profiles and the Swedish coast, or that the megashear does not extend so far south.

We interpret the northeast-dipping reflectors that penetrate into the mantle (F in Fig. 3b) as being related to the northeast-dipping subduction zone previously inferred^{7,16,17} as the cause of formation of the 1.89–1.86 Gyr arc terranes and granitoids in this region. The reflectors at the base of the crust (C) that continue into the upper mantle (E–F) may represent a remnant of Proterozoic oceanic crust trapped in a subduction zone choked by collision of arc or continental terranes from the south with the Skellefte arc.

We interpret the northeast-dipping crustal reflectors to represent an along-strike continuation of the northeast-dipping conductive body detected by magnetotelluric measurements in Sweden³⁴. The conductive and reflective body may be an accretionary prism that was thickened in the collision zone, either containing highly graphitic and sulphidic metasedimentary rocks (analogous to the exposed Flambeau conductive anomaly in the north central United States⁴⁶ associated with the 1.85-Gyr Niagara suture zone¹ or containing water trapped during Proterozoic subduction (analogous to dipping conductors above the active Vancouver Island subduction zone in western Canada⁴⁷).

FIG. 4 Profile of Fig. 3, recorded by a vertical-component seismometer located at the star in Fig. 2, with travel times reduced at 8 km s^{-1} , filtered from 5 to 20 Hz, and with automatic gain control applied with a 1.25-s window. Reflections M1 and M2 indicated by arrowheads correspond to reflectors at depths and locations shown by M1 and M2 in Fig. 3b. Inset shows one-dimensional velocity–depth function above M1 interpreted from these data.

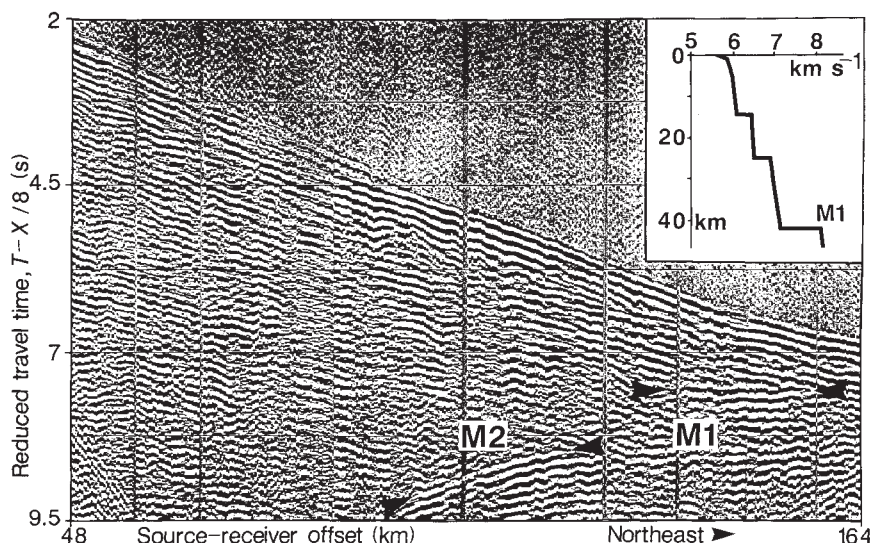
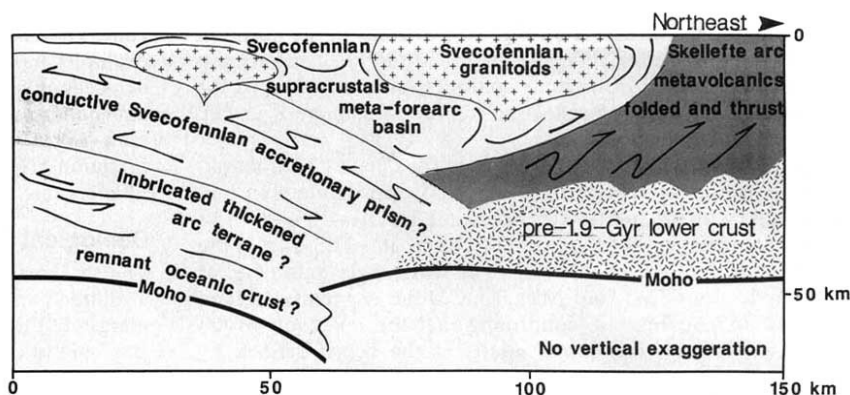


FIG. 5 A diagrammatic interpretation of data in Fig. 3 to show present-day crustal structure resulting from a collision at ~1.86 Gyr between an oceanic Svecofennian arc from the south with the ~1.89-Gyr continental Skellefte arc to the north.



Our interpretation is a uniformitarian plate tectonic model (Fig. 5) consistent with both the seismic data (Figs 3 and 4) and the surface geology (Fig. 2). The most important difference from Phanerozoic orogens is the crustal thickness of 42–48 km, in sharp contrast to western Europe where, away from the Cenozoic and Mesozoic mountain chains in which crustal thickness is still evolving, Palaeozoic mountain chains have crustal thicknesses of only 30–35 km (ref. 48). It has been uncertain whether the greater thickness of Proterozoic crust compared with Phanerozoic crust is due to differences in the way it was formed⁴⁹ or due to continual gradual growth of the continents by igneous intrusion⁵⁰. The preservation beneath Bothnian Bay of a vertical Moho offset of perhaps 15 km implies that the lower crust and uppermost mantle in this region have not been strongly heated or deformed⁵¹ since the formation of this Moho offset at >1.8 Gyr BP. Neither the post-orogenic, pre-1.8-Gyr uplift necessary to expose amphibolite-facies rocks¹⁴, nor the continued intrusion of late Svecofennian granites until 1.78 Gyr (ref. 22), nor the intrusion of dolerite sheets in the region of the Bothnian Sea at 1.25 Gyr (ref. 28), can have produced sufficient heating or deformation at the base of the crust to erase the >1.8-Gyr-old crustal root. This suggests that the present crustal thickness is comparable with the original Proterozoic crustal thickness and has not been significantly increased by subsequent magmatic underplating or intraplating.

We note that the lithospheric thickness in this central part of the Baltic shield (>170 km)⁵² is also much greater than lithospheric thickness beneath adjacent Phanerozoic terranes (<110 km), although our data do not constrain the age of this lithosphere.

Conclusions

New deep-reflection profiling across a presumed early Proterozoic suture zone in the Baltic shield shows a Moho offset and crustal reflector geometries which we attribute to collisional tectonics during a plate tectonic cycle. Our interpretation of early Proterozoic plate tectonics is supported by regional geological mapping and by comparison of the reflector geometries with those observed in Phanerozoic collisional belts. Although other reflection profiles have been recorded across possible Proterozoic sutures⁵³, our data provide the clearest seismic evidence yet for Proterozoic plate tectonics. The kinematics of plate collision seem not to have changed significantly over 1.9 Gyr, even though the 1.93–1.86 Gyr orogenies in the Baltic shield apparently involved crust 40–50 km thick, substantially thicker than is preserved in many Phanerozoic orogenic belts. Not only do our data strongly support uniformitarian tectonic models, they also imply that lower-crustal and intramantle reflectors can survive, and preserve a geometric record of ancient orogeny for over 1.85 Gyr. □

Received 18 June; accepted 12 September 1990.

- Hoffman, P. F. in *The Geology of North America* Vol. A (eds Bally, A. W. & Palmer, A. R.) 447–512 (Geol. Soc. Am., Boulder, 1989).
- Nakajima, T. et al. *Nature* **346**, 263–265 (1990).
- Warner, M. R. *Am. geophys. Un., Geodyn. Ser.* **13**, 281–286 (1986).
- Klemperer, S. L. in *Digital Seismology and Fine Modelling of the Lithosphere* (eds Cassinis, R., Nolet, G. & Panza, G. F.) 228–258 (Plenum, New York, 1989).
- Berthelsen, A. in *Proc. 1st Workshop on the European Geotraverse, the northern segment* (eds Galson, D. A. & Mueller, St.) 13–22 (Euro. Sci. Found., Strasbourg, 1984).
- Gaál, G. & Gorbatschev, R. *Precamb. Res.* **35**, 15–52 (1987).
- Gaál, G. *Bull. geol. Soc. Finland* **58**, 149–168 (1986).
- Skiöld, T. & Öhlander, B. *Precamb. Res.* **45**, 19–26 (1989).
- Öhlander, B., Skiöld, T. & Nisica, D. H. (abstr.) *2nd Symp. on the Baltic Shield Abstr. Vol.*, 69 (Lund Univ., Lund, 1990).
- Huhma, H. *Geol. Surv. Finland Bull.* **337**, 1–48 (1986).
- Kontinen, A. *Precamb. Res.* **35**, 313–341 (1987).
- Park, A. F. *Geology* **13**, 725–729 (1985).
- Lundberg, B. *Geol. För. Stockh. Föhr.* **102**, 156–166 (1980).
- Gaál, G. *Precamb. Res.* **46**, 83–114 (1990).
- Vassjoki, M. & Vivallo, W. *Spec. Pap. geol. Surv. Finland* **10**, 39–40 (1989).
- Hietanen, A. *J. Res. U.S. Geol. Surv.* **3**, 631–645 (1975).
- Berthelsen, A. in *The Anatomy of Mountain Ranges* (eds Schaer, J.-P. & Rodgers, J.) 31–57 (Princeton Univ. Press, 1987).
- Witschard, F. *Precamb. Res.* **23**, 273–315 (1984).
- Welin, E. *Precamb. Res.* **35**, 95–113 (1987).
- Skiöld, T. *Precamb. Res.* **38**, 147–164 (1988).
- Einarrson, Ö. *Sver. geol. Unders. Afh.* **C748**, 1–123 (1979).
- Patchett, P. J., Todd, W. & Gorbatschev, R. *Precamb. Res.* **35**, 145–160 (1987).
- Claesson, S., Huhma, H., Kinny, P. & Williams, I. S. (abstr.) *2nd Symp. on the Baltic Shield Abstr. Vol.*, 27 (Lund Univ., Lund, 1990).
- Korsman, K. (ed.) *Bull. geol. Surv. Finland* **343**, 1–96 (1988).
- Berthelsen, A. & Marker, M. *Tectonophysics* **128**, 163–181 (1986).
- Åxberg, S. *Stockh. Contrib. Geol.* **36**, 151–213 (1981).
- Flöden, T., Jacobsson, R., Kumpas, M. G., Wadstein, P. & Wannäs, K. *Geol. För. Stockh. Föhr.* **101**, 321–327 (1980).
- Gorbatschev, R., Solyom, Z. & Johansson, I. *Geol. För. Stockh. Föhr.* **101**, 177–190 (1979).
- Luosto, U. & Korhonen, H. *Tectonophysics* **128**, 183–208 (1986).
- Luosto, U. in *Proc. 6th Workshop of the European Geotraverse, data compilations* (eds Freeman, R. & Mueller, St.) 53–63 (Eur. Sci. Found. Strasbourg, 1990).

- Lund, C.-E. in *Proc. 6th Workshop of the European Geotraverse* (eds Freeman, R. & Mueller, St.) 65–70 (Eur. Sci. Found., Strasbourg, 1990).
- Guggisberg, B. & Berthelsen, A. *Terra Cognita* **7**, 631–638 (1987).
- Balling, N. in *Proc. 1st Workshop on the European Geotraverse* (eds Galson, D. A. & Mueller, St.) 53–68 (Eur. Sci. Found., Strasbourg, 1984).
- Rasmussen, T. M., Roberts, R. G. & Pedersen, L. B. *Geophys. J. R. astr. Soc.* **89**, 799–820 (1987).
- Hjelt, S.-E. in *Proc. 6th Workshop of the European Geotraverse* (eds Freeman, R. & Mueller, St.) 287–297 (Euro. Sci. Found., Strasbourg, 1990).
- Korja, T. thesis, Oulu Univ. (1990).
- Klemperer, S. L., Hauge, T. A., Hauser, E. C., Oliver, J. E. & Potter, C. J. *Geol. Soc. Am. Bull.* **97**, 603–618 (1986).
- Mooney, W. & Brocher, T. *Rev. Geophys.* **25**, 723–742 (1987).
- Frei, W. et al. *Nature* **340**, 544–547 (1989).
- Choukroun, P. & ECORS Team *Tectonics* **8**, 23–39 (1989).
- Phinney, R. A. *Am. geophys. Un., Geodyn. Ser.* **14**, 157–172 (1986).
- Freeman, B., Klemperer, S. L. & Hobbs, R. W. *J. geol. Soc. Lond.* **145**, 727–740 (1988).
- Klemperer, S. L., Ryan, P. D. & Snyder, D. B. *J. geol. Soc. Lond.* (in the press).
- Elming, S.-Å., Thunehed, H., Lindqvist, G. & Palm, H. *Geol. För. Stockh. Föhr.* (in the press).
- Henkel, H. *Svenska Kärnbränslehantering AB Tech. Rep. Vol.* 88-07, 1–66 (Stockholm, Sweden, 1987).
- Sternberg, B. K. & Clay, C. S. *Am. geophys. Un., Geophys. Monogr.* **20**, 501–530 (1977).
- Hyndman, R. D. *J. geophys. Res.* **93**, 13391–13405 (1988).
- Meissner, R., Wever, Th. & Flügel, E. *Ann. Geophys.* **65**, 357–364 (1987).
- Meissner, R. *The Continental Crust 1–426* (Academic, London, 1986).
- Nelson, K. D. *Geophys. J. Int.* (in the press).
- Kuznir, N. J. & Matthews, D. H. *J. Petrol. Spec. Lithosphere Iss.*, 63–87 (1988).
- Calagnile, G. *Tectonophysics* **90**, 19–35 (1982).
- Cannon, W. F., Lee, M. W., Hinze, W. J., Schulz, K. J. & Green, A. G. *Geology* (in the press).
- Boyd, R. et al. *Bull. geol. Surv. Finland* **333**, map (1985).

ACKNOWLEDGEMENTS. We thank the seismic and marine crews of the S/V *Mintrop* for their help, E. Jarvis for processing the data shown in Fig. 3a, and R. Gorbatschev and A. Hirn for reviews. All the reflection data are being processed by Prakla-Seismos A G and will become available from the British Geological Survey. Project BABEL was funded by NERC, the Commission of the European Community (Project JOULE), BIRPS Industrial Associates (Amerada Hess, BP, Britoil, Chevron, Conoco, Mobil, Shell), the Swedish Natural Science Research Council, the Academy of Finland, the Deutsche Forschungsgemeinschaft, and the Danish Natural Science Research Council. S.L.K. was supported by the Royal Society, Cambridge Earth Sciences Contribution number 1767.

Invariant chain distinguishes between the exogenous and endogenous antigen presentation pathways

Luc Teyton, Deirdre O'Sullivan*, Phillip W. Dickson, Vincent Lotteau, Alessandro Sette*, Pamela Fink[†] & Per A. Peterson

Department of Immunology, Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, California 92037 and * Cytel, 11099 North Torrey Pines Road, La Jolla, California 92037, USA

Class I MHC molecules acquire peptides from endogenously synthesized proteins, whereas class II antigens present peptides derived from extracellular compartment molecules. This dichotomy is due to the fact that the invariant chain associates with class II molecules in the endoplasmic reticulum, preventing binding of endogenous peptides. The mutually exclusive binding of peptide and invariant chain to class II molecules suggests that the invariant chain might play a part in autoimmune disease.

TWO main groups of T lymphocytes recognize foreign antigens in the form of peptides associated with class I and class II major histocompatibility complex (MHC) molecules (reviewed in ref. 1). T cells that serve mainly as helper cells express CD4 and are primarily restricted by class II molecules, whereas the CD8-expressing cells, which mostly represent cytotoxic effector cells, interact with class I molecules². A critical step in immune activation of T cells is the binding of peptides to the MHC molecules. The precise sequence of events leading to the binding is not known, but evidence suggests that class I molecules acquire peptides in the endoplasmic reticulum (ER) and that peptide binding is a prerequisite for the export of the class I antigens out of the ER to the cell surface³. Several lines of evidence suggest that it is in the endosomal compartment that class II molecules associate with peptides generated by proteolytic breakdown of proteins exogenous to the class II-bearing antigen-presenting cells⁴⁻⁶. The reason for this dichotomy has been puzzling in view of the striking structural and functional similarities between class I and class II molecules⁷⁻⁹. Here we show that the invariant chain (I), which is associated with class II antigens intracellularly¹⁰⁻¹², prevents the binding of peptides to class II antigens. Although this observation might explain how endogenous and exogenous antigen presentation pathways are kept apart, it raises questions about presentation of self-peptides in autoimmune diseases.

Isolation of a secreted I chain

It has been shown that I chains associate with class II molecules in the ER¹¹⁻¹⁶, but the biological consequences of this interaction remain obscure. The early suggestions that the I chain influences the intracellular transport and/or post-translational modifications of class II molecules have been questioned¹⁷⁻¹⁹. Likewise, a role for the I chain in regulating the endocytosis of class II antigens²⁰ has not been confirmed²¹. To further address the question of whether the I chain might block peptide binding

to class II molecules^{22,23}, we have generated a secreted, soluble form of the molecule. A schematic representation of the I chain, (Fig. 1a) shows that it is a type II transmembrane protein. The hydrophobic sequence serves both as signal sequence and membrane anchor²⁴. We modified a full-length Iip31 human complementary DNA²⁵ such that the region encoding the N-terminal 29 residues was removed. The resulting construct should unmask the signal sequence cleavage site and lead to the production of a secreted soluble form of the I chain (see ref. 25). The modified cDNA was transferred into a baculovirus expression vector²⁶. Cells infected with the recombinant virus were biosynthetically labelled and chased for various periods of time. A protein reactive with a rabbit antiserum against the I chain precipitated a molecule of the expected molecular weight from the culture medium (Fig. 1b). The secretion of this molecule was relatively slow, but the secreted protein was stable and accumulated to concentrations of more than 2 mg litre⁻¹ over a three-day period.

Secreted soluble I chain molecules were isolated from medium of infected cells. The isolated I chains were homogeneous by size exclusion chromatography and two-dimensional gel electrophoresis (Fig. 1c). Separate analyses demonstrated that the purified, soluble I chain occurred as a dimer. It contained O-linked and N-linked carbohydrate moieties. The N-linked sugars were terminally glycosylated, as judged by resistance to endonuclease H digestion.

Binding of I chains to class II

The I chain binds all isotypes of MHC class II molecules both in man^{10,27} and mouse¹⁶. The interaction involves the extra-cytoplasmic region of the I chain²⁸.

The highly purified, secreted I chain molecules were used to measure the interaction between the I chain and human class II antigens. HLA-DR1 and DR7 molecules were separately isolated by immunosorbent chromatography and the highly purified proteins were kept in the presence of 1% N-octyl glucoside. The binding of the radiolabelled class II molecules to I chains displayed concentration dependence and was saturable (Fig. 2a, b). The interaction was rapid and essentially complete within 30 min. The specificity of the interaction was demonstrated by the competition with unlabelled HLA-DR molecules; purified HLA class I molecules did not compete (data not shown). The dissociation constant for the interaction between the DR7 class II molecule and soluble I chain is 1.5×10^{-8} M and 3.0×10^{-9} M, for DR1 (Fig. 2). These values are only approximations as we cannot evaluate what fraction of class II molecules in our preparations is active in terms of binding to the I chain. We show below that I chain and peptides compete for binding to class II molecules, but we cannot conclude that all peptide-associated class II molecules are unable to bind I chain. Peptides may differ in their binding to class II molecules with regard to size and charge and their exact site of binding to class II molecules may vary. Likewise, the three-dimensional structure of the class II molecule may be dependent on the peptide bound. The same assumptions are true when we measure the peptide binding to class II molecules, so we have

[†] Present address: Department of Immunology, University of Washington, Seattle, Washington 98195, USA.

just evaluated the percentage of molecules available in a given preparation. The average binding of peptides to HLA-DR1 was 45% in five separate experiments and 60% for HLA-DR7. The fact that virtually identical values were obtained for the I chain occupancy does not invalidate this reasoning. But, given the fact that a large pool of I chains is present in the ER of cells constitutively expressing class II molecules, the rapid and complete interaction between I chains and class II chains should be ascertained at the estimated dissociation constant (K_d) value for the interaction.

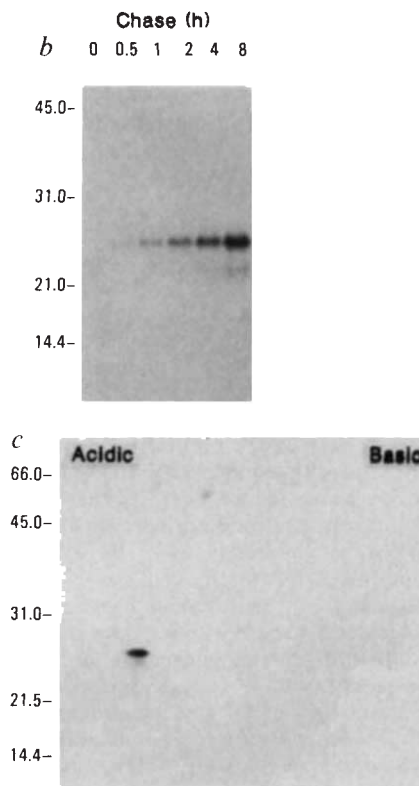
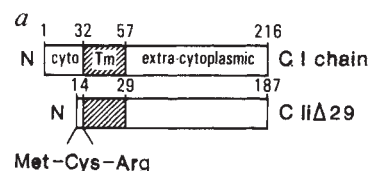
The I chains binds α and β chains

The heterodimeric human class II antigens are comprised of three isotypes^{27,29}, each one of which is composed of a unique

α and β chain²⁹. The class II molecules are highly polymorphic such that each isotype constitutes a separate allelic series. The polymorphic residues are not randomly distributed within the class II molecule but occur as clusters in the first domain of both the α and β chains such as in the class I molecule^{7,8} and these clusters are believed to make up the groove to which peptides bind⁹.

The interaction between class II molecules and the I chain is normally initiated in the ER^{11,16}. Pulse-chase experiments in lymphoblastoid cell lines have shown that in the ER, free α and β chains have a half life of only a few minutes before they are assembled into oligomeric structures¹¹. To investigate whether I chains preferentially interact with class II antigen α or β chains, HeLa cells—which do not express I chains or any of

FIG. 1 *a*, Schematic representation of full length human lip31 and the truncated mutant li Δ 29 where the cytoplasmic tail has been deleted and replaced by the sequence Met-Cys-Arg. *b*, Accumulation in the medium of soluble I chain from *Spodoptera Frugiperda* cells infected with a recombinant baculovirus containing the li Δ 29 cDNA. *c*, Purity of isolated, ¹²⁵I-labelled, soluble I chain by two-dimensional gel electrophoresis. M_r (K) on the left. METHODS. The full-length cDNA clone p γ -2 (ref. 25) was subcloned into the PstI site of the pBS vector (Stratagene). Complementary oligonucleotides 5'-CGGATCCATTATGTGCCGC-3' and 5'-GGCACATAATGGA-3' were synthesized using standard cyanoethyl phosphoramidite chemistry, kinased and annealed to generate a linker that reconstructs an ATG translation-initiation codon and creates a BamHI cloning site 5' to it, an SphI cloning site at the 5' end and a SacII cloning site at the 3' end. Linkers were ligated into SphI-SacII cut and purified p γ -2 pBS plasmid. The construct was sequenced using the double-stranded DNA as a template. SacII cuts right at the boundary of the cytoplasmic and transmembrane regions. The first three amino acids encoded by li Δ 29 are Met, Cys, and Arg and are followed by the normal transmembrane stretch, which is used as a signal sequence by the translational machinery²⁴. The modified cDNA retained with a BamHI fragment was first subcloned into pCMU4 and checked for expression in HeLa cells³⁰. The BamHI fragment was then inserted into the BamHI cloning site of the pAC373 vector²⁶. Plasmids containing a single copy of li Δ 29 in the correct orientation were identified by restriction mapping. Transfer of the li Δ 29 sequence from pAC373-li Δ 29 to the AcNPV genome was as described²⁶. The pAC373-li Δ 29 plasmid (2.5 μ g) was combined with 1 μ g of AcNPV DNA and transfected by calcium-phosphate precipitation into *Spodoptera Frugiperda* cells (SF9). Homologous recombination between the polyhedrin sequence of pAC373 and the wild-type virus generates recombinants that no longer form inclusions in infected cells. Plaque-purified recombinant viruses were tested by Southern blot with a ³²P-labelled I chain cDNA probe to verify the integration of the li Δ 29 sequence. High-titre viral stocks were generated by infecting SF9 cells grown in Grace's insect medium (Gibco) supplemented with 0.33% Difco T C yeastolate, 0.33% lactalbumin hydrolysate (Gibco), 2 mM glutamine, and 10% FCS (Irvine Scientific, California) at 1 multiplicity of infection (MOI). The cells were cultured at 1×10^6 ml⁻¹ for 4 days. We characterized the li Δ 29 secretion from SF9 cells by [³⁵S]methionine labelling of infected SF9 cells. In brief, 1×10^6 SF9 cells were seeded per well in 6-well plates and infected at a MOI of 10 in 0.2 ml. The inoculum was removed and cells were cultured in 1 ml of Grace's medium at 27 °C for 2 days. Cells were then washed twice in 1 ml methionine-deficient Grace's medium and incubated in the same medium for 1 h. The cells were then labelled with 100 μ Ci of [³⁵S]methionine per well for 15 min and chased for different times (see figure). Culture supernatants were collected, and centrifuged at 4 °C for 30 min at 100,000 g in a Beckman LT-55 ultracentrifuge. The supernatants were pre-cleared overnight at 4 °C with 5 μ l of rabbit serum and 100 μ l of Pansorbin (Calbiochem), microfuged for 30 min. Immunoprecipitations were carried out with K2, a rabbit antiserum raised against the C-terminal peptide KESLEEDPSSGLGVTKQSL (single-letter amino acid code) of the I chain (V. Quaranta, unpublished data), for 2 h at 4 °C and followed by protein A-Sepharose treatment for 20 min. The beads were washed three times with 0.5% NP-40 in PBS and three times with 0.5% NP-40 in 20 mM Tris buffer, pH 7.6, before elution with endoglycosidase H (Endo H) buffer (0.1% sodium dodecylsulphate, 0.05 M sodium acetate, pH 5.6) at 95 °C for 3 min. Samples were run in 12% acrylamide SDS-polyacrylamide gels. For large-scale production of li Δ 29, secondary virus stocks were generated in serum-free ExCell-400™ medium (J. R. Scientific, California) and used to infect SF9 cells grown in the same medium at 1×10^6 ml⁻¹ and cultured for 3 days at 27 °C in spinner flasks. At this stage, supernatants were collected by centrifugation and supplemented with protease inhibitors: 1 mM phenylmethylsulphonyl fluoride, 1 μ g ml⁻¹ leupeptin, 1 μ g ml⁻¹ pepstatin, 1 μ g ml⁻¹ chymostatin and 0.05 μ g ml⁻¹ trypsin



inhibitor. After dialysis against 50 mM Tris, pH 7.6, the material was applied to a DEAE Sepharose column (Pharmacia). Bound proteins were eluted with a NaCl gradient (0–300 mM). Samples of each fraction were tested for the presence of soluble li Δ 29 by western blotting⁴² using the K2 antiserum. li Δ 29 was eluted between 180 and 220 mM NaCl. The fractions containing I chain were pooled, concentrated to 4 ml on an ultrafiltration cell (model 8010, Amicon) and dialysed against 150 mM NaCl, 50 mM Tris, pH 7.6 before chromatography on a Sephacryl S-200 superfine column (Pharmacia). Fractions containing I chain were identified as before, pooled, dialysed against 50 mM Tris, pH 7.6, and chromatographed over a MonoQ column (Pharmacia) with a bed volume of 8 ml. li Δ 29 emerged from the MonoQ column as a homogeneous peak. Finally, this material was run over a Superose 6 column after concentration and dialysis. The fractions containing the purified li Δ 29 were pooled, concentrated to 1 mg ml⁻¹ and kept at –80 °C. The purity was checked by two-dimensional gel electrophoresis¹⁰ of ¹²⁵I-labelled material using NEPHGE in the first dimension and 12% SDS-PAGE in the second dimension. li Δ 29 was labelled by the chloramine T method as described previously⁴³. The spot appearing at 67 K on the two-dimensional gel was traces of albumin labelled during the radioiodination procedure when the column buffer containing BSA was mixed with the sample.

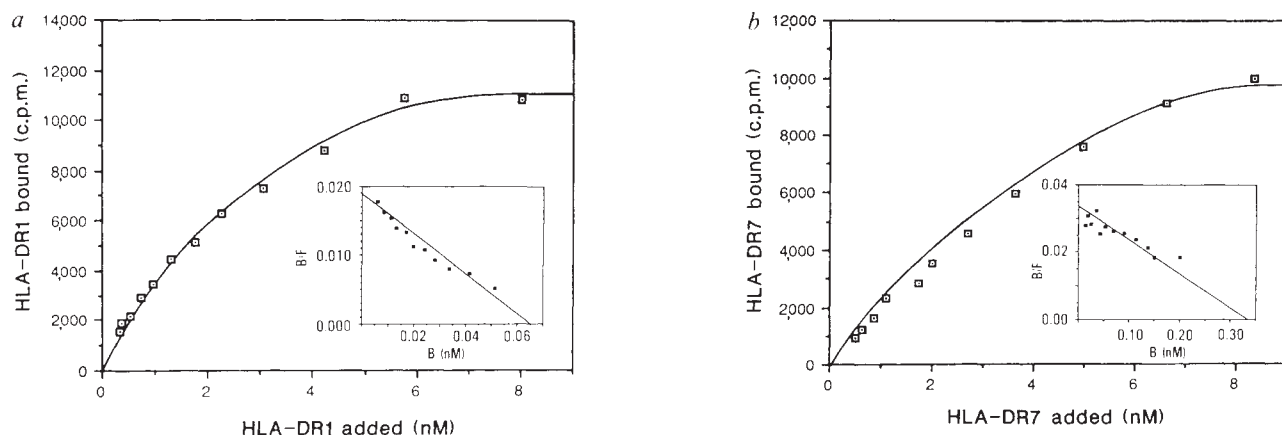


FIG. 2 Binding of HLA-DR1 (a) and HLA-DR7 (b) class II molecules to soluble invariant chain. The inserts show the corresponding Scatchard analyses⁴⁴. METHODS. HLA-DR molecules were affinity-purified using the monoclonal antibody LB 3.1 from EBV-transformed homozygous cell lines essentially as described³⁵. The purified DR molecules were radioiodinated at room temperature with immobilized lactoperoxidase-glucose oxidase reagents (Enzymobeads, BioRad) for 10 min in 250 mM sodium phosphate, pH 7.5, 0.4% β -D glucose, 0.5 mCi ¹²⁵Iodine. Free iodine was separated from labelled HLA-DR molecules on a PD-10 column equilibrated with 1% N-octyl glucoside,

1 mg ml⁻¹ BSA, 1 mM PMSF. Increasing concentrations of ¹²⁵I-labelled HLA-DR molecules (0.5–10 nM) were incubated with 70 nM of purified I Δ 29 for 2 h at room temperature in a final volume of 20 μ l in the presence of 1% N-octyl glucoside and the protease inhibitor cocktail. The HLA-DR-I Δ 29 complexes were separated from free HLA-DR molecules by immunoprecipitation with a saturating concentration of the K2 antiserum for 2 h, followed by protein-A Sepharose adsorption. Each point represents the mean of duplicates.

the class II α and β chains—were transiently transfected with cDNAs³⁰ encoding DR α , Dr β and Iip31 chains, respectively, cloned into the expression vector pCMU-4 (ref. 30). Individual chains or combinations of two chains were expressed. Radio-labelled proteins were immunoprecipitated with antisera specific for α , β and I chains and analysed by SDS-PAGE. Figure 3 summarizes the results. Iip31 alone gave rise to two labelled proteins with relative molecular masses of 31,000 and 25,000 (M_r of 31K and 25K). The 31K I-chain molecule represents the full-length protein, whereas the smaller polypeptide is a proteolytic fragment corresponding to the extracytoplasmic part of the I chain³¹ and which arises by cleavage of the full-length I chain in the ER (unpublished data). The DR β chain alone gave rise to a single radioactive band. Co-expression of the I chain with the individual DR subunits demonstrated that the I chain can independently bind both α and β chains. Endonuclease H digestions confirmed that none of the interacting chains had undergone terminal glycosylation. These results confirm that the I chain interacts with β chains³², but show that the α chain also has affinity for the I chain. The fact that the I chain can interact with individual class II chains suggests that it either has two independent sites for binding of the α and β chains or a single site that recognizes similar structural features of the two subunits. We cannot yet distinguish between these possibilities, but it seems reasonable to assume that the I chain forms complexes with the individual class II chains in the ER before the α - β assembly. A possible role for the I chain in the class II antigen assembly process must, therefore, be entertained, even if it has

been shown that the α - β heterodimer can form in the absence of I chain^{18,19}.

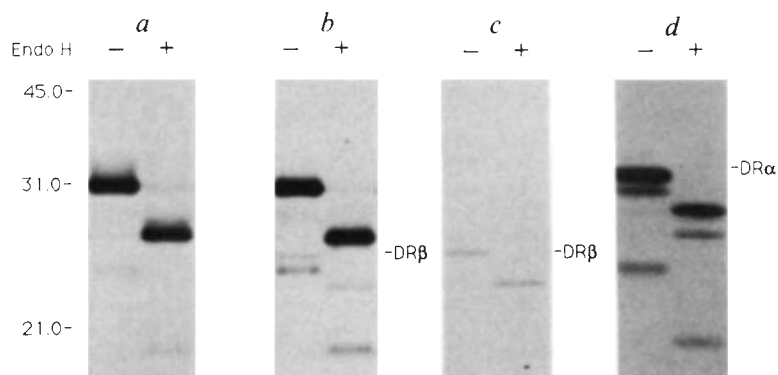
I chain affects the peptide-binding groove

If the I chain interacts with regions of the individual class II subunits that display structural similarities, then it may bind to the non-polymorphic parts of the class II molecule. Only those portions of the α and β chains comprising the immunoglobulin-like, membrane-proximal domains (α 2 and β 2) display a striking sequence homology²⁹. A similar tertiary structure for those domains seems likely. But a proposed model for the three-dimensional structure of class II antigens⁹ based on the recently solved crystallographic structure of class I antigens^{7,8} opens the possibility that the two membrane-distal α and β domains (α 1 and β 1)—which form the peptide-binding groove and encompass the polymorphic residues of the molecule—may be folded similarly, to generate a pseudo-symmetrical structure built on a backbone of conserved residues. Thus, structural considerations alone do not allow obvious conclusions to be drawn as to the most likely site of interaction between the I chain and the two class II chains.

To address this issue, we used monoclonal antibodies directed against the two domains of the DQ β chain. One antibody, CA.206 (ref. 10), recognizes all class II antigen β chains and is directed against the β 2 domain (unpublished data). The other antibody, GSP 200.1, distinguishes DQ α 3.1 from DQ α 3.2 such that it only reacts with the latter molecule³³. Sequence analyses have shown that these two molecules differ by only six amino

FIG. 3 Interaction of I chain with individual α and β HLA-DR chains. Cells transfected with cDNAs encoding I chain (A, B and D), HLA-DR β chains (B and C) and HLA-DR α chains (D) were labelled with [³⁵S]methionine and immunoprecipitations of I chain (A and B), HLA-DR β chains (C), and HLA-DR α chains (D) were carried out before SDS-PAGE. Half of each immunoprecipitate was treated with Endo-H. M_r (K) markers on the left hand side.

METHODS. HeLa cells were transfected with the indicated cDNAs cloned into the expression vector pCMU4 using the calcium-phosphate method³⁰. Then, 72 h after the transfection, the cells were labelled with 0.2 mCi of [³⁵S]methionine per 1×10^7 cells for 15 min. Labelled proteins were immunoprecipitated, Endo H-treated and analysed by SDS-PAGE as described³⁰.



acid residues, four of which occur in positions 13, 26, 45 and 57 of the $\beta 1$ domain. Residue 45 is critical for the expression of the epitope recognized by GSP 200.1 (ref. 33) which suggests that this antibody binds close to the peptide-binding groove of the DQw3.2 molecule.

The effect of the two monoclonal antibodies on the binding of purified DQw3.2 to the I chain was examined (Fig. 4). CA2.06 did not measurably impede the interaction between the DQw3.2 molecules and the I chain and had no more effect than the control antibody OKT8. In contrast, GSP 200.1, which binds to the $\beta 1$ domain of the DQw3.2 molecule inhibited the interaction between the class II molecule and the I chain in a concentration-dependent manner. The antibody IV.D.12, which is DQw3 reactive³⁴ and which is likely to recognize the $\beta 1$ domain, had the same inhibitory effect as GSP 200.1 (not shown). These data suggest either that the interaction between the I chain and assembled class II antigens only involves the β chain, or that the antibodies, for steric reasons, inhibit the binding of I chain both to α and β chains. In any event, this observation indicates that the I chain preferentially interacts with the membrane-distal domain of class II antigen β chains.

The I chain prevents peptide binding

To examine whether the binding of I chain to class II antigens would affect the interaction with peptide, we incubated separately two allelic forms of class II antigens with the radiolabelled peptides corresponding to residues 307–319 of the influenza hemagglutinin (peptide 520.3) and 830–843 of the tetanus toxoid (peptide 512) in the presence of various concentrations of I chain. Previous analyses have shown that peptides 515 and 520.3, which differ by a tyrosyl residue, bind to DR1 molecules, whereas peptide 512 binds to DR2 molecules³⁵. The I chain efficiently prevented the peptides from binding to the class II antigens (Fig. 5A), which confirms and extends recent data²³. On a molar basis the I chain seemed to be as effective as the unlabelled peptides in preventing association between the radiolabelled peptides and the class II antigens. As the binding of I chain to the DR molecule is rapid (half-maximal binding occurs within 15 min at room temperature), whereas the *in vitro* binding of peptides to class II antigens is slow (equilibrium between free and bound peptide requires more than 24 h at room temperature)²², these data do not distinguish between the I chain acting as a competitive or a non-competitive inhibitor of peptide binding. Preliminary analyses suggest that I chain has little effect on the off-rate of the peptide–class II interaction, but the precise mechanism of competition between the I chain and peptides for binding to class II molecules is not known.

Two extreme situations can be envisaged with regard to the

inhibitory effect of the I chain on the peptide–class II molecule interaction. As I chain binds more rapidly than peptides do, it is conceivable that it sterically blocks the binding groove of the class II molecules such that peptides are denied access. If this were true, soluble I chain might interact with class II antigens regardless of whether they contain bound peptides or not. Alternatively, soluble I chain may bind directly to the peptide-binding groove of class II molecules, such that binding of I chain and peptides is mutually exclusive. The latter situation would also arise if I-chain binding caused a conformational change in the class II molecule such that the peptide-binding groove would become functionally impaired.

To investigate the nature of the peptide modulated interaction between I chain and class II molecules, we incubated class II DR1 molecules with the radiolabelled influenza hemagglutinin peptide 307–319 and a limiting amount of soluble I chain to obtain ~50% maximal binding of the peptide. As expected, gel chromatography separations demonstrated that the added I chain prevented a significant fraction of the labelled peptide from associating with the class II molecules (Fig. 5B, a and b). The elution position of the radiolabelled peptide bound to class II molecules was the same regardless of whether the I chain was present or not, suggesting that the I chain was not bound to those class II molecules that contained the peptide. This was corroborated by directly monitoring the distributions of the class II molecules and I chain, respectively, in the column effluent. The class II molecules emerged as two peaks in the presence of I chain (Fig. 5B). The material comprising the later eluted peak occurred coincidentally with that of the radiolabelled peptide. This elution position is identical to that obtained for class II molecules chromatographed in the absence of I chain or peptide. Separate analyses of the distribution of the soluble I chain demonstrated that this protein was also distributed into two peaks (Fig. 5B). The first eluted peak coincided with the highest molecular mass form of the class II molecules (compare Fig. 5B, c and d) and, therefore, represents complexes between I chains and class II molecules. The second peak was eluted slightly later than that for peptide-containing class II molecules. That elution position is identical to the one recorded for soluble I chains chromatographed in the absence of class II antigens. These analyses suggest that the binding of I chains and peptides to class II antigens might be mutually exclusive, but the different kinetics of each interaction limit that interpretation. The gel chromatography profiles indicate that it is the dimeric form of the soluble I chain that binds to class II molecules. Whether both subunits of the I chain molecule are engaged in the interaction remains to be established. The availability of mutant class II chains might allow us to address this issue in the future.

FIG. 4 Antibody inhibition of the class II molecule–I chain interaction. The binding of purified HLA-DQw3.2 molecules to purified, soluble I chain was impeded by the antibody GSP 200.1 (◆) but not by CA 2.06 (■) and OKT8 (□). The results are expressed as percentages of maximal binding.

METHODS. Nunc Maxisorp 96-well plates were coated overnight at 4 °C with 50 μ l per well of a 2 μ g ml⁻¹ solution of soluble I chain in PBS. After washing with 0.05% Tween-20 in 150 mM NaCl, 50 mM Tris, pH 7.6, (TBS) the plates were blocked with 0.25% gelatin, and 0.05% Tween-20 in TBS for 2 h at room temperature. GSP 200.1 (ref. 33) and CA 2.06 (ref. 10) were titrated by ELISA onto coated plates containing affinity purified HLA-DQw3.2 molecules. DQw3.2 molecules at a final concentration of 10 μ g ml⁻¹ were incubated in duplicate for 2 h at room temperature in the presence or absence of antibodies in 0.1% Tween-20-TBS. OKT8 antibodies (Becton-Dickinson) were used as a control. The binding of HLA-DQw3.2 was determined after successive 1 h incubations with a rabbit serum specific for the HLA-DQ α cytoplasmic tail diluted 1:5,000 and a peroxidase-conjugated goat anti-rabbit IgG diluted 1:7,500 (BioRad). Each step was followed by five washes with 0.1% Tween-20-TBS. After addition of 100 μ l per well of a solution containing the substrate OPD, the optical density at 492 nm was measured in a BioRad microplate reader.

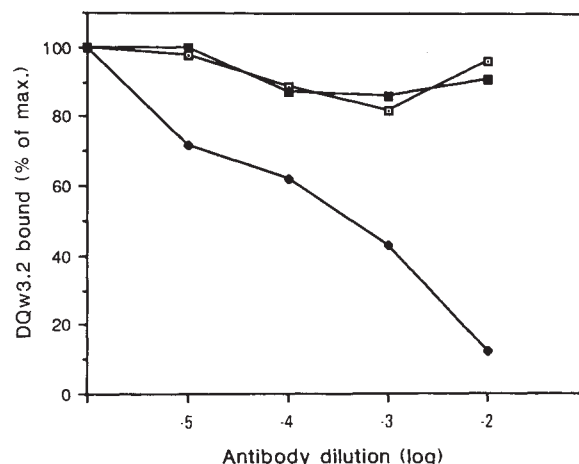
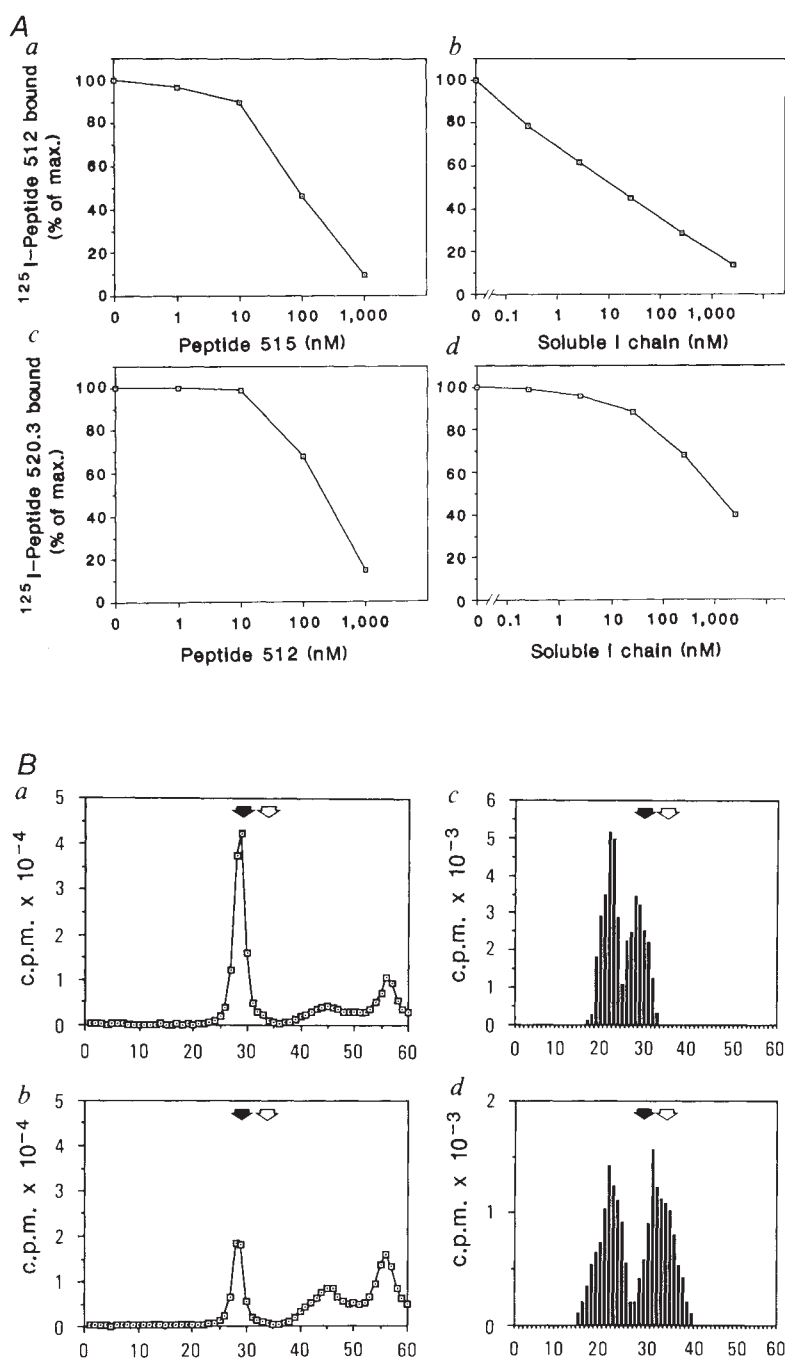


FIG. 5 A, Soluble I chain inhibits peptide binding to class II antigens. *a*, Inhibition of binding of labelled 520.3 peptide to HLA-DR1 molecules by unlabelled 515 peptide. *b*, Inhibition of binding of 520.3 peptide to HLA-DR1 molecules by soluble I chain. The binding of labelled 520.3 to HLA-DR1 in the absence of inhibitor was 48.0%. *c* and *d*, The inhibition of binding of labelled 512 peptide to HLA-DR2 molecules with unlabelled 512 (*c*) and soluble I chain (*d*), respectively. The binding of labelled 512 peptide to HLA-DR2 molecules in the absence of inhibitor was 45.4%. *B*, Samples containing ^{125}I -labelled 520.3 peptide and HLA-DR1 molecules were separated on a Superose 6 chromatography column with (*b*) or without (*a*) prior addition of soluble I chain. *c* and *d*, Elution profiles of HLA-DR1 and I chain molecules of the sample shown in (*b*). Open arrow, the position of free I chain. Closed arrow, the position where free HLA-DR molecules were eluted. **METHODS.** The HLA-DR-peptide binding was assayed as previously described³⁵. Purified HLA-DR1 (0.35 μM) was incubated with 10 nM ^{125}I -labelled 520.3 peptide for 48 h in the presence of the protease inhibitor cocktail. Peptide 515 is the hemagglutinin peptide 307–319. Peptide 520.3 is the same but a tyrosine residue has been added at the N-terminus to allow iodination. Purified HLA-DR2 molecules were incubated using the same conditions with radiolabelled 512 peptide (tetanus toxin peptide 830–843). Purified peptides were labelled using the chloramine T method⁴³. HLA-DR-peptide complexes were separated from free peptides by gel filtration on a Sephadex G50 column, as previously described^{22,35}. The inhibitors, unlabelled peptides or soluble I chain, were added at the beginning of the reaction. In *A* the results are expressed as the percentage of maximal binding. *B*, Purified HLA-DR1 molecules at a final concentration of 0.35 μM were incubated with 10 nM ^{125}I -labelled 520.3 peptide with or without 0.3 μM of soluble I chain (54.8% inhibition of binding; maximal binding of 520.3, 45.1%). The samples were separated by gel chromatography on a Superose 6 column and 0.5 ml fractions were collected. Radioactivity profiles were determined. HLA-DR and I chain distributions were evaluated using a solid phase radioimmunoassay performed on Immobilon P filters (Millipore). Aliquots of each column fraction (50 μl) were spotted in duplicate onto the membrane using a 96-well sample manifold. The membrane was allowed to dry for 1 h at 37 °C and was then treated for 1 h with 10% monoethanolamine in 1.0 M sodium bicarbonate buffer, pH 9.5, for capping. Antibody binding was performed in 0.1% Tween-20, 3% BSA in TBS (binding buffer) for 2 h at room temperature with purified IgG from a rabbit antiserum specific for DR α chain, and purified IgG from antiserum K2 for the detection of I chain. The filters were washed five times with 0.1% Tween-20-TBS and then incubated for 1 h at room temperature with 1:2,000 dilution of biotinylated anti-rabbit IgG (Amersham) in binding buffer. After five washes, the filters were incubated with 1×10^6 c.p.m. ml^{-1} of radiolabelled streptavidin (Amersham) for 20 min in binding buffer. After additional washing, bound radioactivity was measured. Background binding was less than 200 c.p.m. for both assays.



The I chain prevents antigen presentation

Only small amounts of I chain are normally present on the surface of class II antigen-expressing cells³⁶. Higher concentrations of I chain might impede the ability of class II molecules to present peptides to T lymphocytes. To address this issue directly, we used a well characterized T-cell hybridoma recognizing the moth cytochrome *c* peptide 88–103 in the context of I-E^k. The IL-2 secretion of this hybridoma was stimulated in a dose-dependent manner by the presence of the appropriate peptide and antigen-presenting cells (Fig. 6a). Fixed cells can also present the antigen, demonstrating that the peptide does not need processing to be recognized by the T cells.

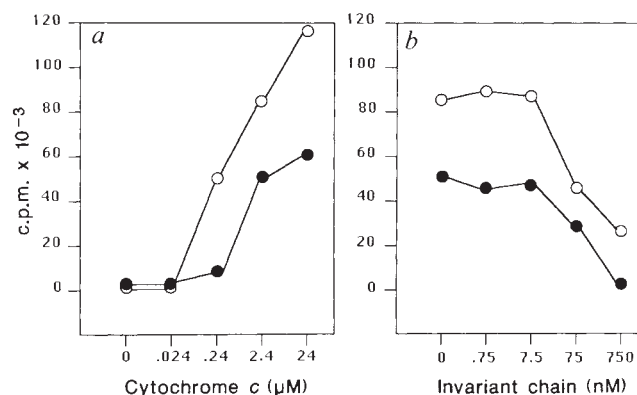
When the antigen-presenting cells were incubated with the peptide and increasing concentrations of soluble I chain, the activation of the T cells measured as IL-2 secretion was gradually impaired. Activation of T-cells by concanavalin A was not affected by I chains (not shown). The inhibitory activity of I chains was quantitatively the same with fixed and unfixed anti-

gen-presenting cells suggesting that the effect of I chains was independent of endocytosis. Separate experiments demonstrated that soluble human I chain binds both to I-A and I-E molecules (not shown). Separate experiments also demonstrated that the soluble human I chain bound to the antigen-presenting cells but not to the T cells. We conclude that the inhibitory effect of the I chain was due to it directly interfering with peptide binding to class II molecules on the surface of the antigen-presenting cells.

Discussion

These data indicate an important function for I chains in the physiology of MHC class II molecules and the CD4-restricted antigen presentation process. Thus, when the I chain interacts with class II molecules in the ER, it prevents endogenous peptides from binding to these molecules²³. The efficiency of the system might depend more on quantitative differences between the two types of interactions than on I chains always

FIG. 6 Soluble I chain inhibits antigen-dependent, class II-restricted IL-2 production by T-cell clone HOD6 in a dose-dependent manner. **a**, Dose-dependent response of the HOD6 T cell clone to moth cytochrome *c* peptide 88-103 presented by differently treated B10.A antigen-presenting cells (APC): ○, freshly irradiated cells, ●, freshly fixed and irradiated cells. **b**, Inhibition of IL-2 production by HOD6 cells in the presence of 2.4 μ M of the moth cytochrome *c* peptide and increasing concentrations of soluble I chain. METHODS. HOD6 is an I-E^k-restricted, cytochrome *c*-specific, CD4⁺ T-cell clone of B10.S(9R) origin⁴⁵. HOD6 cells were grown in DMEM and 10% FCS. For the antigen stimulation, 1×10^5 HOD6 cells were incubated with 5×10^5 irradiated (3,000 rads) B10.A APCs and increasing concentrations of moth cytochrome *c* peptide 88-103 in a total volume of 200 μ l in flat bottom 96-well plates. Cell-free supernatants were removed from duplicate wells at 24 h and added at a final concentration of 33% to 5×10^3 IL-2-dependent HT-2 cells that had been washed twice in Hank's balanced salt solution (HBSS) to remove any contaminating IL-2. These cells were incubated for one day, pulsed with 1 μ Ci per well of [³H]thymidine and collected 18 h later. Counts incorporated into high *M*_r material were measured in a scintillation counter. For pulsing experiments, APCs were incubated at 1×10^7 cells ml⁻¹ in complete medium in the presence of 2.4 μ M of moth cytochrome *c* peptide 88-103 for 90 min at 37 °C, with frequent mixing. Cells were washed twice in HBSS before fixing. The fixing was done after the removal



of red blood cells from the APC suspension by osmotic lysis. The cells were then washed and incubated at 5×10^6 ml⁻¹ in HBSS with 0.05% glutaraldehyde for 30 s at room temperature. An equal volume of 0.2 M L-lysine was added to stop the reaction. The cells were washed and resuspended in medium. I chain was diluted in HBSS.

outcompeting peptides from binding to nascent class II molecules. Indeed, in the ER the I chain has two advantages over any peptide: its abundance relative to the level of class II molecules, and its faster kinetics of interaction relative to the kinetics of the peptide-class II molecule interaction. Consequently, we assume that peptide binding to class II antigens is unlikely to occur in the ER.

Class I antigens seem to bind peptides that are confined to intracellular compartments so the immune recognition of pathogens with an intracellular life cycle is ascertained. As class II antigens divest themselves of the I chain in the endosomal compartment³⁷⁻³⁹, antigenic peptides derived from exogenous proteins would not have to displace endogenous peptides to bind to class II molecules. In contrast, all class I molecules seem to have bound peptides before they reach the cell surface³ and this should severely impede the binding of peptides derived from the exogenous pathway. The I chain might, therefore, be the single most important factor in distinguishing the endogenous from the exogenous antigen presentation pathway.

The exact sites of binding of I chain on the class II molecules remain obscure. It seems likely that the I chain has at least one

site of interaction on each class II chain and that the site on the β chain resides in the β -1 domain. This finding does not imply that the I chain interacts directly with any of the polymorphic residues present in the domain. It is possible that the interaction is indirectly affected, however, by residues in the peptide-binding groove. Should this be the case, it seems likely that the I chain would display varying affinities for different isotypic and allelic forms of class II molecules. Provided situations exist when the amount of I chain is limiting, for example, under conditions of lymphokine-induced expression of class II molecules, then class II molecules with relatively weak affinities for I chain might not be protected from binding endogenous peptides. Thus, the well documented association between certain allelic forms of class II molecules and autoimmune diseases⁴⁰ may depend less on the particular endogenous peptides presented but more on how efficiently the I chain prevents the presentation of such peptides. The intriguing observation that Asp 57 in DQ β chains gives protection from insulin-dependent diabetes⁴¹ may mean that Asp 57 strengthens the interaction with the I chain such that endogenous peptides are more efficiently excluded than with other DQ β chains. □

Received 18 June; accepted 20 September 1990.

1. Townsend, A. & Bodmer, H. A. *Rev. Immun.* **7**, 601-624 (1989).
2. Swain, S. L. *Immunol. Rev.* **74**, 129-142 (1983).
3. Townsend, A., Ohlen, C., Bastin, J., Foster, L. & Karre, K. *Nature* **340**, 443-448 (1989).
4. Ulanue, E. R. & Allen, P. M. *Science* **236**, 551-557 (1987).
5. Ziegler, K. & Ulanue, E. R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 175-178 (1982).
6. Guagliardi, L. E. *et al. Nature* **343**, 133-139 (1990).
7. Bjorkman, P. *et al. Nature* **329**, 506-512 (1987).
8. Bjorkman, P. *et al. Nature* **329**, 512-519 (1987).
9. Brown, J. H. *et al. Nature* **332**, 845-850 (1988).
10. Charron, D. & McDevitt, H. O. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6567-6571 (1979).
11. Kvist, S., Wiman, K., Claesson, L., Peterson, P. A., & Dobberstein, B. *Cell* **29**, 61-69 (1982).
12. Machamer, C. E. & Cresswell, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1287-1291 (1984).
13. Nguyen, Q. W. & Humphreys, R. E. *J. biol. Chem.* **264**, 1631-1637 (1989).
14. Sung, E. & Jones, P. P. *Molec. Immun.* **18**, 899-913 (1981).
15. Quaranta, V. *et al. J. Immun.* **132**, 1990-1905 (1984).
16. Jones, P. P., Murphy, D. B., Hewgill, D. & McDevitt, H. O. *Immunochimistry* **16**, 51-60 (1978).
17. Claesson-Welsh, L. & Peterson, P. A. *J. Immun.* **135**, 3551-3557 (1985).
18. Sekaly, R. P., Tonnelle, C., Strubin, M., Mach, B. & Long, E. O. *J. exp. Med.* **164**, 1490-1504 (1986).
19. Miller, J. & Germain, R. N. *J. exp. Med.* **164**, 1478-1489 (1986).
20. Stockinger, B. *et al. Cell* **56**, 683-689 (1989).
21. Peterson, M. & Miller, J. *Nature* **345**, 172-174 (1990).
22. Buus, S., Sette, A. & Grey, H. M. *Immunol. Rev.* **98**, 115-141 (1987).
23. Roche, P. A. & Cresswell, P. *Nature* **345**, 615-618 (1990).
24. Lipp, J. & Dobberstein, B. *Cell* **46**, 1103-1112 (1986).
25. Claesson, L., Larhammar, D., Rask, L. & Peterson, P. A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7395-7399 (1983).

26. Kuroda, K., Hauser, C., Rott, R., Klenk, H. D. & Doerfler, W. *EMBO J.* **5**, 1359-1365 (1986).
27. Carra, G. & Accolla, R. S. *J. exp. Med.* **165**, 47-63 (1987).
28. Marks, M. S. & Cresswell, P. *J. Immun.* **136**, 2519-2525 (1986).
29. Kaufman, J. F., Auffray, C., Korman, A. J., Shackelford, D. A. & Strominger, J. *Cell* **36**, 1-13 (1984).
30. Nilsson, T., Jackson, M. & Peterson, P. A. *Cell* **58**, 707-718 (1989).
31. Thomas, L. J., Nguyen, Q. V., Elliott, W. L. & Humphreys, R. E. *J. Immun.* **140**, 2670-2674 (1988).
32. Buerstedde, J. M. *et al. J. exp. Med.* **168**, 823-837 (1988).
33. Kwok, W. W., Lotshaw, C., Milner, E. C. B., Knitter-Jack, N. & Nepom, G. T. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1027-1030 (1989).
34. Giles, R. C. *et al. J. exp. Med.* **157**, 1461-1470 (1983).
35. O'Sullivan, D. *et al. J. Immun.* in the press.
36. Wraight, C. J. *et al. J. biol. Chem.* **265**, 5787-5792 (1990).
37. Nowell, J. & Quaranta, V. *J. exp. Med.* **162**, 1371-1376 (1985).
38. Blum, J. S. & Cresswell, P. *Proc. natn. Acad. Sci. U.S.A.* **85**, 3975-3979 (1988).
39. Nguyen, Q. V., Knapp, W. & Humphreys, R. E. *Hum. Immun.* **24**, 153-163 (1989).
40. Sveigaard, A., Platz, P. & Ryder, L. P. *Immunol. Rev.* **70**, 193-218 (1983).
41. Todd, J. A., Bell, J. I. & McDevitt, H. O. *Nature* **329**, 599-604 (1987).
42. Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350-4354 (1979).
43. Greenwood, F. C., Hunter, W. M. & Glover, J. S. *Biochem. J.* **89**, 114-123 (1963).
44. Scatchard, G. *Ann. N.Y. Acad. Sci.* **51**, 660-672 (1949).
45. McElligott, D. L., Sorger, S. B., Matis, L. A. & Hedrick, S. M. *J. Immun.* **140**, 4123-4131 (1988).

ACKNOWLEDGEMENTS. We thank V. Quaranta and S. Radka for antibodies; L. Matis and O. Kanagawa for the HOD6 hybridoma; J. Sidney and P. Turel for technical assistance; and D. Ulanowski for secretarial assistance. P.W.D. is a special fellow of The Leukemia Society of America. P.J.F. is a scholar of The Leukemia Society of America. V.L. is supported by the Diabetes Research Foundation. This work was supported by the NIH (P.J.F. and P.A.P.).

New structures near the compact radio source at the Galactic Centre

F. Yusef-Zadeh*, M. Morris† & R. D. Ekers‡

* Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA

† Department of Astronomy, University of California, Los Angeles, California 90024, USA

‡ Australia Telescope National Facility, PO Box 76, Epping, NSW 2121, Australia.

SIXTEEN years ago, Balick and Brown¹ found a bright compact radio source, Sgr A*, at the Galactic Centre, as had been predicted by Lynden-Bell and Rees². The characteristics of its emission and the kinematics of gas near it have been used to argue for the presence of a black hole of $500\text{--}5 \times 10^6$ solar masses at the Galactic Centre^{3–5}. Here we present high-resolution (<0.2 arcsec) radio images, made at 2-cm wavelength, of the central arcsecond of the Galaxy. Sgr A* is surrounded by at least four satellite radio features. Interpreted as thermally emitting blobs of plasma, they have typical sizes $<1,700$ AU and are within a projected distance of 3,500 AU from Sgr A*. Another set of larger radio features, 2–4 arcsec from Sgr A*, shows a regular progression of galactocentric distance with position angle. The placement and appearance of all these structures lead us to suggest that they are physically associated with Sgr A*; if so, they can be used as probes of the immediate vicinity of the possible massive black hole at the Galactic Centre.

The non-thermal radio source Sgr A* is the most intense radio object in the inner several degrees of the Galactic Centre, with a brightness temperature T_b exceeding 7×10^8 K at 1.3-cm wavelength ($\lambda = 1.3$ cm) and a radio luminosity of $\sim 2 \times 10^{34}$ erg s⁻¹. Its flux density varies with time by $\leq 20\%$, and the frequency dependence of its spectrum is $\nu^{1/3}$, consistent with

that expected of an optically thick, compact non-thermal radio source^{6–9}. The characteristics of Sgr A*, unique in our Galaxy, resemble those of non-thermal point sources in the cores of extragalactic radio sources in galactic nuclei, but Sgr A* is much less luminous. Because of these similarities, and because of the kinematics of the gas near Sgr A*, some have considered it to be a candidate for a massive black hole, although its mass has been the subject of debate^{3–5}.

We made radio continuum observations at $\lambda = 2$ cm in March 1986, November 1987, December 1986 and October 1988 using the VLA in all four of its configurations A, B, C and D, respectively. The visibility data were calibrated in all observing sessions by using 3C286 and NRAO530 as the amplitude and phase calibrators. 3C286 was assumed to have a flux density of 3.443 Jy at 14.965 GHz and 3.450 Jy at 14.915 GHz. Maximum u, v (lengths of baselines in projected units of wavelength) coverage was obtained during each 8-h observing run. The phase centre was placed about $1''$ northeast of Sgr A*. A standard self-calibration procedure was applied several times to the data sets from the different array configurations before and after they were combined.

The key to making high-dynamic-range images was to place the position of the maximum flux density of Sgr A* exactly at the centre of a grid cell before using the deconvolution algorithms. This allowed the CLEAN deconvolution algorithm to derive a proper model of the visibility data, and consequently, the self-calibration procedure was effective in minimizing the antenna-based phase and amplitude errors. The highest-resolution images described here have a dynamic range $\geq 1.5 \times 10^4$ and are constructed by combining data from only the most expanded configurations (A and B arrays), with a maximum baseline of 36 km.

By using a maximum-entropy (ME) algorithm following the removal of 1.153 Jy from the central position of Sgr A*, we were able to achieve a spatial resolution approaching $0.07'' \times 0.14''$ (the pixel size is $0.035'' \times 0.035''$, and the full width at half maximum (FWHM) of the CLEAN beam is $\sim 0.12'' \times 0.24''$, with a position angle PA of 7.8°) and to achieve an r.m.s. noise of

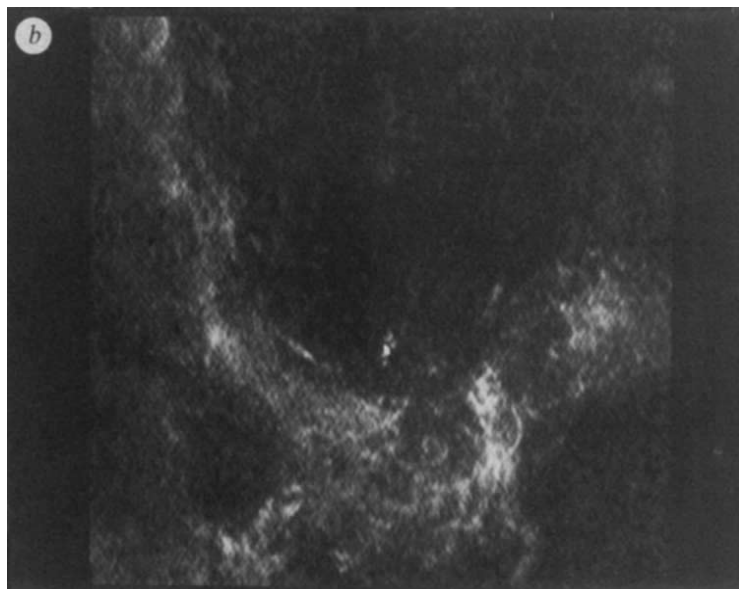
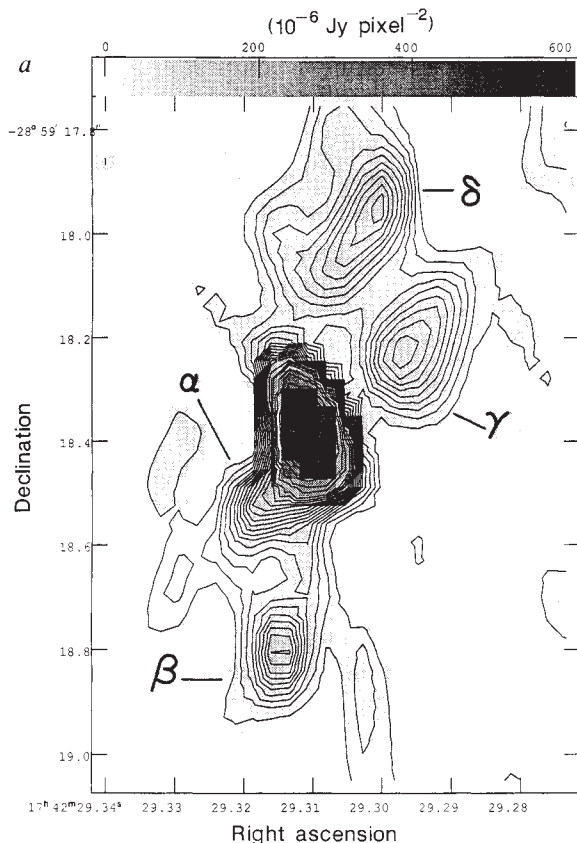


FIG. 1 *a*, The 2-cm ME brightness distribution of Sgr A* with a spatial resolution of $0.12'' \times 0.24''$ (PA = 7.8°). Contours are 2, 3, 4, ..., 14, 16, 18, 20, 25, 30, 40, 50, 60, 80, 100, 130, 160×10^{-6} Jy pixel⁻². *b*, The radiograph shows the inner light year of the Galactic Centre. The peak emission from features α to η , the northern arm of Sgr A West, and a segment of the 'bar' of ionized gas in which a small cavity is embedded are visible.

75 μ Jy per beam area, which is about twice the theoretical thermal noise. From a detailed comparison of Fig. 1 with other images generated by using a CLEAN algorithm instead of an ME algorithm, we conclude with confidence that the new substructures in Fig. 1 are not artefacts of the deconvolution process and that the beam pattern, which is symmetric about its centre, has been removed quite effectively from the final images. Furthermore, an inspection of 2-cm images constructed from data taken using two different array configurations (A and B) shows similar and consistent results for the new features presented here. Because Sgr A* is the only bright source in the field, we cannot rule out the possibility that combined residual phase and amplitude calibration errors may be responsible for producing sources α to δ (see below). Although the more extended ϵ to η plasma blobs (see below) have been detected at a number of frequencies, sources α to δ need to be confirmed.

Figure 1a shows four radio structures in the vicinity of Sgr A*. These features are located on opposite sides of the source along two axes, one making an angle of about 50° and the other lying at about 75° relative to the galactic plane. The orientation and placement of these new features indicate that they are physically associated with the source at the Galactic Centre. In a $3.2'' \times 3.2''$ box centred on Sgr A* (the box size is chosen to avoid emission in the radio 'bar'), there is only one other source brighter than 0.07 mJy pixel $^{-2}$ except for source ϵ , which, as we argue below, may be associated with Sgr A*. Furthermore, in the relatively blank area to the north of Sgr A*, there are essentially no sources of comparable intensity except for a few associated with known infrared sources, such as IRS7. Therefore, the areal density of sources is small enough that the cluster centred on Sgr A* is unlikely to have occurred by chance superposition of foreground and background sources. Nevertheless we cannot rule out the possibility that these sources are unrelated to Sgr A* until further observations are carried out. The northernmost blobs have angular diameters in the range $\sim 0.2''$ – $0.4''$ and have an integrated flux density of ~ 1 – 3 mJy and a peak intensity of $\sim 10^{-4}$ Jy pixel $^{-2}$, corresponding to $T_b \sim$

500 K at $\lambda = 2$ cm. The integrated flux density of the four nearby subcomponents is ~ 10 mJy, which is $<1\%$ of the total flux density of Sgr A*. We note that features γ and δ appear to be pointing approximately toward Sgr A* (right ascension (1950) = $17^h 42^m 29.31^s$, declination (1950) = $-28^\circ 59' 18.38''$) and approximately perpendicular to the galactic plane (position angle of about 30°).

Examining a much larger region than shown in Fig. 1a, we recently discussed an apparent cavity⁹ in the ionized 'bar' of Sgr A West¹⁰. This 'mini-cavity' has a diameter of $\sim 2''$ and its centre is offset from Sgr A* by $\sim 3''$. The mechanism responsible for creating this cavity has not been clearly identified. The new images shown in Figs 1b and 2 with two different spatial resolutions reveal a chain of approximately circular features—labelled ϵ , ξ and η —which appear to form a spatial progression through the cavity, a progression that might originate at Sgr A* (lower-resolution 6-cm images, such as Fig. 1 of ref. 7, also depict the relationship of sources ϵ , ξ and η to the mini-cavity and possibly to Sgr A*). Although we cannot rule out the possibility that the placement of these radio peaks is entirely fortuitous, we suggest that the coincidence be pursued. It is interesting in this regard that observations of Ne II emission towards the cavity indicate that the gas kinematics there deviate strongly from circular rotation about the centre, with ionized gas having velocities that may be as large as ~ 400 km s $^{-1}$ (ref. 3 and J. Lacy, personal communication). Inasmuch as the radio peaks in the progression correspond to high-velocity thermal plasma blobs, they may have displaced the ionized matter in the 'bar' of Sgr A West, thereby creating the mini-cavity (a more detailed account of this interpretation will be given elsewhere).

Sources α to δ may be ionized lumps in a very intermittent wind emanating from Sgr A*. Although this interpretation is not unique, it supports the suggestion that a powerful wind from Sgr A* is blowing away the modest mass loss from a supergiant star, IRS7, which lies near the Galactic Centre (F.Y.-Z. and M.M., manuscript in preparation). The assumption that Sgr A* is surrounded by ionized gas is not inconsistent with the fact

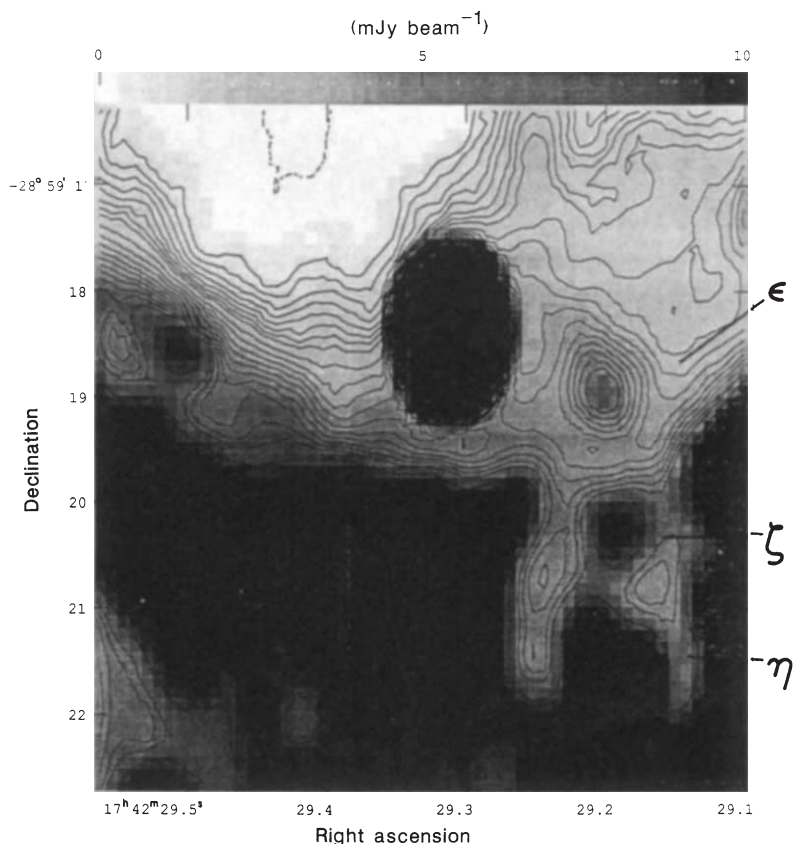


FIG. 2 The CLEAN image formed from tapered visibility data at $\lambda = 2$ cm with FWHM = $0.39'' \times 0.61''$ (PA = -0.7°) with contours of total intensity at $-3, 3, 4, 5, \dots, 10, 12, 14, 16, 18, 20$ mJy per beam area.

that the interstellar medium in the inner two parsecs of the Galaxy is dominated by ionized gas constituting the structure of Sgr A West. Assuming that the blobs are thermally emitting, optically thin homogeneous spheres with an electron temperature of 10^4 K and a path length of $\sim 8 \times 10^{-3}$ pc (0.2"), and using the emission measure of source δ (5×10^7 cm $^{-6}$ pc), we estimate its electron density to be 9×10^4 cm $^{-3}$. The total ionized mass of an individual blob is $M \approx 8 \times 10^{-4} M_{\odot}$.

The blob velocities are of particular interest. If one assumes that the $4 \times 10^6 M_{\odot}$ believed to be lurking in the inner 0.1 pc is concentrated in Sgr A*, then the orbital velocity of a blob located 2,500 AU from Sgr A* is $\sim 1,200$ km s $^{-1}$. This is similar to the velocity required of the intervening plasma inhomogeneities if the time variability of Sgr A* is due to the motion of these scattering centres¹¹. The orbital velocity is a minimum velocity, because both outflow and infall velocities would be larger unless the blobs all happened to be caught when they are near their maximum radial extent. The dynamical timescale for the blobs α to δ would therefore be about 10 years. If Sgr A* is very massive, the association of blobs α to δ with it would be verifiable in the near future by direct observation of their spatial displacements and/or their line-of-sight velocity. Any observed displacement would presumably reveal whether these features have been ejected from the neighbourhood of the putative black hole at the Galactic Centre or whether they represent accretion onto it. If there is no noticeable motion of the blobs on a 10-yr timescale, then we should look elsewhere for the concentrated mass at the Galactic Centre. □

Received 11 April; accepted 19 August 1990.

1. Balick, B. & Brown, R. L. *Astrophys. J.* **194**, 265–270 (1974).
2. Lynden-Bell, D. & Rees, M. J. *Mon. Not. R. astr. Soc.* **152**, 461–475 (1971).
3. Serabyn, E., Lacy, J. H., Townes, C. H. & Bharat, R. *Astrophys. J.* **326**, 171–185 (1988).
4. Ozernoy, L. M. *IAU Symp. 136, The Center of the Galaxy* (ed. Morris, M.) 555–566 (Kluwer, Dordrecht, 1989).
5. Allen, D. A. & Sanders, R. H. *Nature* **319**, 191–194 (1986).
6. Lo, K. Y., Cohen, M. H., Readhead, A. C. S. & Backer, D. C. *Astrophys. J.* **249**, 504–512 (1981).
7. Lo, K. Y. *IAU Symp. 136, The Center of the Galaxy* (ed. Morris, M.) 527–534 (Kluwer, Dordrecht, 1989).
8. Jauncey, D. L. *et al. Astr. J.* **98**, 44–48 (1989).
9. Yusef-Zadeh, F., Morris, M. & Ekers, R. D. *IAU Symp. 136, The Center of the Galaxy* (ed. Morris, M.) 443–451 (Kluwer, Dordrecht, 1989).
10. Ekers, R. D., van Gorkom, J. H., Schwartz, U. J., Downes, D. & Rogstad, D. J. *Astr. Astrophys.* **122**, 143–150 (1985).
11. Zhao, J.-H., Ekers, R. D., Gross, M. W. & Lo, K. Y. *IAU Symp. 136, The Center of the Galaxy* (ed. Morris, M.) 535–541 (Kluwer, Dordrecht, 1989).

ACKNOWLEDGEMENTS. We thank T. Cornwell for discussions. This research was supported by the office of the Vice President for Research at Northwestern University. The VLA is operated by Associated Universities Inc., under agreement with the NSF. M.M. acknowledges support from the NSF.

Carbon isotope composition of individual amino acids in the Murchison meteorite

M. H. Engel*, S. A. Macko† & J. A. Silfer*

* School of Geology and Geophysics, 100 E Boyd Street, University of Oklahoma, Norman, Oklahoma 73019, USA

† Department of Environmental Sciences, University of Virginia, Charlottesville, Virginia 22903, USA

A SIGNIFICANT portion of prebiotic organic matter on the early Earth may have been introduced by carbonaceous asteroids and comets¹. The distribution and stable-isotope composition of individual organic compounds in carbonaceous meteorites, which are thought to be derived from asteroidal parent bodies, may therefore provide important information concerning mechanistic pathways for prebiotic synthesis² and the composition of organic matter on Earth before living systems developed³. Previous studies^{11,12} have shown that meteorite amino acids are enriched in

¹³C relative to their terrestrial counterparts, but individual species were not distinguished. Here we report the ¹³C contents of individual amino acids in the Murchison meteorite. The amino acids are enriched in ¹³C, indicating an extraterrestrial origin. Alanine is not racemic, and the ¹³C enrichment of its D- and L-enantiomers implies that the excess of the L-enantiomer is indigenous rather than terrestrial contamination, suggesting that optically active materials were present in the early Solar System before life began.

Hypotheses concerning extraterrestrial prebiotic synthesis of organic compounds are often based on the composition of organic matter in carbonaceous meteorites^{4,5}. Attention continues to be focused on the organic-rich Murchison meteorite (Type CM, $\sim 2\%$ total organic carbon), which was collected shortly after impact in Australia in 1969. Compared to terrestrial materials, the overall distribution and relative abundance of amino acids in Murchison meteorite stones is exotic, as evidenced by high concentrations of non-protein amino acids that are uncommon in living systems (such as α -aminoisobutyric acid and isovaline) and the apparent absence or extremely low concentration of many of the amino acids common in proteins (such as serine, phenylalanine, tyrosine, lysine, histidine and arginine)^{4,6}.

Several amino acids that are common in living systems also occur in fairly high abundance in the Murchison meteorite (glycine, alanine, glutamic acid, aspartic acid and proline, for example)^{4,7}. The expectation has been that these amino acids, when found in meteorites, will be racemic because laboratory simulations of prebiotic synthesis result in the formation of racemic amino acids⁸. Our previous finding⁶ that some of these amino acids are not racemic (L-enantiomer predominating) led to speculation on whether a percentage of the L-enantiomers of these amino acids reflects terrestrial contamination or whether diagenetic reactions occurring in space explain this anomaly. So far, no one has found a terrestrial contaminant that could rapidly and selectively introduce into a meteorite stone small quantities of some of the protein amino acids (such as glycine, alanine, aspartic acid and glutamic acid), but not others (such as phenylalanine, serine, tyrosine and lysine)⁷.

Pillinger⁹ has suggested that the stable-isotope composition of the D- and L-enantiomers of the individual amino acids might indicate whether non-racemic amino acids were indigenous to the meteorite or a consequence of contamination. This idea is based on the observation that terrestrial organic matter is depleted in ¹³C (ref. 10) and that bulk extracts of the amino acids in carbonaceous meteorites are enriched in ¹³C (refs 11, 12). A significant amount of an L-amino-acid contaminant in a meteorite stone should therefore result in a marked stable-isotope depletion for this component. Because the concentrations of amino acids in the Murchison meteorite are of the order of nanomoles per gram, previously available methods^{13–15} would require the destruction of large quantities of sample to isolate individual enantiomers for stable-isotope analysis. The recent development of a method for amino acid analysis that combines gas chromatography and isotope-ratio mass spectrometry (GC-IRMS), however, allows direct measurement of the stable-carbon-isotope composition of individual enantiomers in a water extract of a Murchison stone.

We isolated samples weighing 50 mg and 3.8 g from the interior of a 52.3-g Murchison stone (R. A. Langheinrich meteorite collection) that had a largely intact fusion crust. Portions of the 50-mg sample were acidified (6 N HCl), dried and then prepared for stable carbon and nitrogen isotope analysis by the Dumas sealed-tube method¹⁷. Analyses of the resultant gases were performed with a VG PRISM isotope-ratio mass spectrometer. Other portions of the bulk, untreated sample of the meteorite were analysed directly for stable-carbon-isotope composition and elemental abundance (C, N) using a VG Isogas Carlo Erba combustion system, and for stable-isotope (C, O) composition of the carbonate component using the VG PRISM.

The 3.8-g sample was ground to a fine powder and placed in a Pyrex tube. Deionized, distilled H₂O (20 ml) was added to the tube, which was sealed under N₂ and heated for 8 h in an oven at 100 °C. After heating, the water extract and a procedural blank were desalted and prepared for amino acid derivatization as previously reported⁶. Unlike previous studies^{6,12}, we did not hydrolyse the water extract with 6 N HCl before desalting. Although hydrolysis increases the recovery of amino acids (presumably by the alteration of labile precursor¹⁸), we decided to try to obtain stable-isotope compositions for the water-extractable amino acids alone, as the addition of amino acids formed by hydrolysis might complicate interpretation of the data. We determined the abundances of amino acids in a portion of the water extract using high-performance liquid chromatography (HPLC), as previously reported¹⁹. The remainder of the water-extractable amino acid fraction was derivatized to (N, O)-TFA-isopropyl esters^{16,20} for GC and GC-IRMS analysis using reagents (trifluoroacetic anhydride, isopropanol) of known stable-carbon isotope composition.

The VG Isochrom GC-IRMS system consists of a Hewlett-Packard 5890 GC coupled to a combustion furnace/water trap, which in turn is interfaced to a VGSIRA isotope-ratio mass spectrometer²¹. We used a Chirasil-Val 50 m × 0.25 mm (inner diameter) fused-silica capillary column capable of resolving (N, O)-TFA-isopropyl esters of D and L amino acids²⁰.

Unlike GC-IRMS analyses of hydrocarbons^{22,23}, the analysis of amino acids is complicated by the addition of carbon in the required derivatization. Previous $\delta^{13}\text{C}$ analyses of standard amino acid derivatives by GC-IRMS and conventional static-combustion IRMS were in good agreement¹⁶. For example, the $\delta^{13}\text{C}$ values for N-TFA-isopropyl esters of D-alanine and L-alanine determined by conventional static-combustion IRMS were -29.9 ± 0.1 and $-28.0 \pm 0.1\%$, respectively. The $\delta^{13}\text{C}$ obtained for the same derivatives of D- and L-alanine by the GC-IRMS method were -30.1 ± 0.1 and $-27.5 \pm 0.2\%$, respectively. In general, the error for replicate derivatization and analysis of amino acid standards by either method is better than 0.2‰. The $\delta^{13}\text{C}$ values for amino acids are corrected for the carbon introduced during derivatization¹⁶.

The C and N contents determined for the non-acidified sample of the Murchison stone were 2.2% and 0.27%, respectively. The $\delta^{13}\text{C}$ value for this sample was -9.9% (relative to the PDB standard). These values are in agreement with previously reported values for bulk samples of the Murchison meteorite²⁴. The $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values for the acidified bulk sample of this Murchison stone are -12.3% and $+34.7\%$ relative to PDB and atmospheric N₂, respectively. Stable carbon and oxygen isotope values for the carbonate component of this stone were $+59.9\%$ (PDB) and $+37.8\%$ (standard mean ocean water), respectively. The $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ values for the carbonate are enriched in ^{13}C and ^{18}O relative to previous reports²⁵.

The concentrations of amino acids in the non-hydrolysed water extract of this stone (Table 1) were in good agreement

with a previous analysis of this fraction in another Murchison stone²⁶. Several common protein amino acids were not detected in this sample (phenylalanine, tyrosine, lysine) or were present in trace amounts (serine, threonine, leucine, isoleucine). Again, this is consistent with a previous study of the non-hydrolysed water extract of a Murchison stone²⁶. GC/mass-selective detector analyses confirmed which amino acids were present in this water extract.

The carbon isotope compositions of the major amino acid constituents of the water extract of this Murchison stone are enriched in ^{13}C relative to terrestrial biogenic materials (Table 1), and clearly indicate an extraterrestrial origin. The stable-isotope values for the individual amino acids cover a range of 25‰ and seem to decrease with increasing carbon number, similar to a trend reported by Yuen *et al.*² for hydrocarbons and monocarboxylic acids in the Murchison meteorite. To confirm and interpret this trend with respect to possible kinetic effects during synthesis, however, we must acquire a larger data set.

The D/L values for amino acids in the water extract of this interior stone sample were not racemic and were similar to the D/L values that we previously reported for the hydrolysed water extract of another Murchison stone⁶. Here we find that the D/L value for alanine is 0.85 ± 0.03 . Assuming that the D-enantiomer of alanine is entirely extraterrestrial in origin, the data indicate that the excess L-enantiomer is not from a terrestrial source (Table 1). The L-alanine is 3.0‰ lighter than the D-alanine. If the relative amounts of L-alanine and D-alanine were originally identical, the excess L-alanine required to produce a D/L value of 0.85 would have to have an isotopic composition of $\sim +10\%$ to account for the observed value of $+27\%$ for L-alanine in the water extract. We are not aware of natural terrestrial sources for alanine that are so enriched.

Here and in our previous work alanine was the most racemized amino acid in the water extract. The D/L values for the other amino acids were lower (for example, D/L glutamic acid = 0.54). The $\delta^{13}\text{C}$ value for L-glutamic acid is $+6\%$. Although we cannot, at present, measure the $\delta^{13}\text{C}$ value of D-glutamic acid, we are modifying the GC-IRMS system to allow analysis of the enantiomers (such as D-glutamic acid) that are present in lower abundances.

Our finding of non-racemic amino acids cannot be adequately explained by terrestrial contamination, and may indicate that laboratory simulations of prebiotic syntheses resulting in racemic amino acids are not entirely appropriate models for the chemical reactions that actually took place during formation of the Solar System. Alternatively, extraterrestrial diagenetic reactions after synthesis may have altered the D/L values. The recent finding²⁷ that the Murchison meteorite contains a significant level of ^{14}C formed by cosmic-ray-induced nuclear reactions indicates that organic matter (such as amino acids) in the Murchison meteorite may have been susceptible to extraterrestrial diagenetic reactions, at least since the time of its existence as a smaller body (~ 1 million years) before impact. Although the significance of such reactions and the reactions resulting from β -decay of ^{14}C over the extended residence time in space are unknown, it has been hypothesized that β -decay may have resulted in the preferential destruction of D amino acids on Earth (and thus enrichment in L amino acids) before the origin of life²⁸.

Further refinement of the GC-IRMS method will eventually permit stable-isotope analyses to be performed on other pairs of amino acid enantiomers occurring in lower concentrations in extracts of the Murchison meteorite. Clearly, stable-isotope geochemistry at the molecular level is helping us to understand extraterrestrial, prebiotic organic reactions at about the time of formation of our Solar System. □

TABLE 1 $\delta^{13}\text{C}$ values

Amino acid	Concentration (nmol g ⁻¹)	$\delta^{13}\text{C}^*$ (‰)
α -aminoisobutyric acid	16.0	+5
L-glutamic acid	4.6†	+6
Isovaline	7.5‡	+17
Glycine	28.1	+22
D-alanine	†	+30
L-alanine	12.9†	+27

* The $\delta^{13}\text{C}$ values are corrected for carbon introduced during derivatization¹⁶. $\delta^{13}\text{C}(\text{‰}) = [(^{13}\text{C}/^{12}\text{C})_{\text{sample}} / (^{13}\text{C}/^{12}\text{C})_{\text{PDB}} - 1] \times 10^3$.

† Concentrations reported for L-glutamic acid and L-alanine represent the total contribution of both enantiomers.

‡ This value includes a minor contribution of valine that co-eluted with isovaline during HPLC analysis.

Received 4 May; accepted 12 September 1990.

1. Chyba, C. F., Thomas, P. J., Brookshaw, L. & Sagan, C. *Science* **249**, 366–373 (1990).
2. Yuen, G., Blair, N., Des Marais, D. J. & Chang, S. *Nature* **307**, 252–254 (1984).

3. National Research Council *The Search for Life's Origins* (National Academy Press, Washington, DC, 1990).
4. Cronin, J. R., Pizzarello, S. & Cruikshank, D. P. in *Meteorites and the Early Solar System* (eds Kerridge, J. F. & Matthews, M. S.) 819–857 (University of Arizona Press, 1988).
5. Shock, E. L. & Schulte, M. D. *Geochim. cosmochim. Acta* (in the press).
6. Engel, M. H. & Nagy, B. *Nature* **296**, 837–840 (1982).
7. Engel, M. H. & Nagy, B. *Nature* **301**, 496–497 (1983).
8. Khare, B. N. *et al.* *Icarus* **68**, 176–184 (1986).
9. Pillinger, C. T. *Nature* **296**, 802 (1982).
10. Schidlowski, M. *Nature* **333**, 313–318 (1988).
11. Chang, S., Mack, R. & Lennon, K. *Lunar planet. Sci.* **9**, 157–158 (1978).
12. Epstein, S., Krishnamurthy, R. V., Cronin, J. R., Pizzarello, S. & Yuen, G. U. *Nature* **326**, 477–479 (1987).
13. Engel, M. H. & Macko, S. A. *Analyt. Chem.* **56**, 2598–2600 (1984).
14. Engel, M. H. & Macko, S. A. *Nature* **323**, 531–533 (1986).
15. Serban, A., Engel, M. H. & Macko, S. A. *Org. Geochem.* **13**, 1123–1129 (1988).
16. Silfer, J. A., Engel, M. H., Macko, S. A. & Jumeau, E. J. *Analyt. Chem.* (submitted).
17. Macko, S. A. thesis, Univ. Texas (Austin) (1981).
18. Cronin, J. R. & Moore, C. B. *Science* **172**, 1327–1329 (1971).
19. Hare, P. E., St John, P. A. & Engel, M. H. in *Chemistry and Biochemistry of the Amino Acids* (ed. Barrett, G. C.) 415–425 (Chapman & Hall, London, 1985).
20. Engel, M. H. & Hare, P. E. in *Chemistry and Biochemistry of the Amino Acids* (ed. Barrett, G. C.) 462–479 (Chapman & Hall, London, 1985).
21. Freedman, P. A., Gillyon, E. C. P. & Jumeau, E. J. *Am. Lab.* 114–119 (June 1988).
22. Freeman, K. H., Hayes, J. M., Trendel, J.-M. & Albrecht, P. *Nature* **343**, 254–256 (1990).
23. Kennicutt, M. C. II & Brooks, J. M. *Org. Geochem.* **15**, 193–197 (1990).
24. Kerridge, J. F. *Geochim. cosmochim. Acta* **49**, 1707–1714 (1985).
25. Grady, M. M., Wright, I. P., Swart, P. K. & Pillinger, C. T. *Geochim. cosmochim. Acta* **52**, 2855–2866 (1988).
26. Cronin, J. R. *Origins of Life* **7**, 337–342 (1976).
27. Jull, A. J. T., Donahue, D. J. & Linick, T. W. *Geochim. cosmochim. Acta* **53**, 2095–2100 (1989).
28. Noyes, H. P., Bonner, W. A. & Tomlin, J. A. *Origins of Life* **8**, 21–23 (1977).

ACKNOWLEDGEMENTS. We thank E. C. P. Gillyon and N. Crossley for the use of the instrumentation at VG Isogas. We also thank J. Jumeau for her help with the GC-IRMS analyses and P. Harrigan-Ostrom for her help with the GC/MSD analyses. S.A.M. thanks C. Hillaire-Marcel for providing office and lab space during his sabbatical leave. This work was supported in part by the NSF.

Evidence for stratospheric ozone-depleting heterogeneous chemistry on volcanic aerosols from El Chichón

F. Arnold, Th. Bührke* & S. Qiu*

Max-Planck-Institut für Kernphysik PO Box 103 980,
D-6900 Heidelberg, Germany

STRATOSPHERIC aerosols are thought to promote ozone destruction by reactive halogen compounds in the polar stratosphere. The predominant aerosol usually comprises particles of nitric acid trihydrate; surface-catalysed reactions on these particles produce ozone-destroying halogen compounds and remove reactive nitrogen gases which might otherwise render the halogen compounds inactive^{1,2}. For periods of up to two years after major volcanic eruptions, atmospheric concentrations of sulphuric acid aerosols in the stratospheric eruption cloud may be comparable to those of nitric acid aerosols, and might therefore also play a part in atmospheric heterogeneous chemistry. Indeed, ozone levels in the Northern Hemisphere were observed to decrease following the large eruption of the Mexican volcano El Chichón in 1982^{3,4}. Here we report balloon-borne measurements of trace gases made in June 1982 in the stratospheric eruption cloud of El Chichón. We observe an increase in the concentration of nitric acid, which is probably due to the conversion of reactive nitrogen gases on sulphate aerosols. These reactions are therefore likely to have played a part in the destruction of stratospheric ozone.

Our measurements were made on 10 June 1982 at Gap (44° N; 6° E; southern France) about nine weeks after the eruption of El Chichón, when the stratospheric eruption cloud had reached central Europe (as observed by ground-based lidar at several European stations^{5,6}). The details of the balloon-borne mass spectrometer and preliminary data on gaseous sulphuric acid

have been reported previously⁷. The instrument measures the composition of ambient negative ions, which is sensitively influenced by stratospheric gaseous nitric and sulphuric acids leading to cluster ions of the type $\text{NO}_3^-(\text{HNO}_3)_n$ and $\text{HSO}_4^-(\text{H}_2\text{SO}_4)_l(\text{HNO}_3)_m$. From the measured composition of these ion species the concentrations of ambient gaseous sulphuric acid and nitric acid can be inferred^{8–10}. For sulphuric acid measurements the accuracy and precision are $\pm 50\%$ and $\pm 20\%$, respectively. For nitric acid measurements the accuracy is poor, but the precision is reasonably good ($\pm 20\%$), so we can obtain meaningful information only on the altitude variation of nitric acid.

Figure 1 shows measured absolute concentrations of gaseous sulphuric acid and gaseous nitric acid concentrations. Also given for comparison are typical lidar measurements of the aerosol cloud made in northern Italy⁵, at almost the same latitude as our balloon measurement. The aerosol cloud has a thickness of only one kilometre with a peak around 25.3 km altitude. In the cloud, the concentrations of both gaseous sulphuric and nitric acids were substantially enhanced relative to the concentrations below and above the cloud. The increase of gaseous sulphuric acid was most likely caused by local photochemical production of gaseous sulphuric acid⁷ from SO_2 or other sulphuric-bearing gases in the eruption cloud. Rapid loss of gaseous sulphuric acid occurs by heterogeneous co-condensation with water vapour on pre-existing aerosols resulting in $\text{H}_2\text{SO}_4\text{--H}_2\text{O}$ aerosols. The lifetime for the collision of a gaseous molecule with aerosols is $t_c = 4/VA$ where V is the thermal velocity of the molecule and A is the total surface-area density. For the unperturbed background stratosphere, one obtains $A_0 = 7.5 \times 10^{-9} \text{ cm}^2 \text{ cm}^{-3}$ around 20 km (ref. 3), which, with an ion sticking probability of unity, implies $t_c = 7.4 \text{ h}$. Inside the eruption cloud A was much larger than A_0 and hence t_c must have been much less than 7.4 h. An upper limit for A can be obtained from the total negative-ion concentrations (Fig. 1). In the background stratosphere, the lifetime of negative ions is determined by recombination with positive ions, and is $\sim 1,000 \text{ s}$ at 25 km altitude. Inside a volcanic eruption cloud, A may become sufficiently large to reduce the ambient-ion concentration by ion attachment to aerosols, which would lead to a local minimum in the n_- profile. Our measured n_- profile does not show a significant minimum inside the eruption cloud, which sets an upper limit on A of $2 \times 10^{-7} \text{ cm}^2 \text{ cm}^{-3}$, similar to estimates from aerosol measurements made at middle latitudes³, but 27 times larger than A_0 .

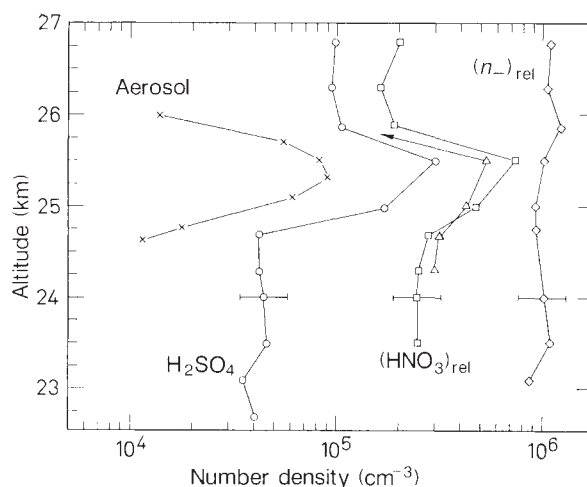


FIG. 1 Gaseous H_2SO_4 (absolute) and HNO_3 (relative) concentrations measured on 10 June 1982 at 44° N, 6° E. HNO_3 concentrations were derived from $\text{HSO}_4^-(\text{HNO}_3)_n$ ($\square$) and $\text{NO}_3^-(\text{HNO}_3)_m$ ($\triangle$). Also given are measured total negative-ion concentrations and aerosol-backscatter ratios (relative units) measured by ground-based lidar⁵. Error bars give precision of measured concentrations.

* Present addresses: VCH-Verlag mbH, D-6940 Weinheim, Germany (Th.B.); and Institute for Space Research, Beijing, China (S.Q.).

There are two possible causes for the observed increase of the gaseous nitric acid concentration (which took place inside the eruption cloud, as proven by sulphuric acid measurements performed in the same air mass and by lidar data): (1) injection of HNO_3 or nitrogen-bearing gases, from which HNO_3 can be formed in the stratosphere; (2) an increased conversion of background reactive nitrogen gases to HNO_3 . Independent observation of a decrease of the total stratospheric NO_2 column after the eruption of El Chichón³ indicates the latter. Also the factor of 2.5 by which the nitric acid concentration increased inside the eruption cloud seems to be roughly consistent with nearly complete conversion of reactive nitrogen to HNO_3 . Model calculations for stratospheric reactive nitrogen gases predict that ~50% of these gases are in the form of nitric acid under unperturbed summer conditions at 25 km and at middle latitudes. Hence, a nearly complete conversion to nitric acid would result in an increase in the nitric acid concentration by a factor of about two. Potential causes for the conversion of reactive nitrogen gases to nitric acid are an increase of the hydroxyl radical concentration and the surface-catalysed reaction



where the subscript *c* denotes condensed phase H_2O contained in H_2SO_4 - H_2O aerosols. The former may be due to the injection of water vapour by the volcanic eruption, followed by OH production from H_2O . But OH also destroys nitric acid and is an important HNO_3 sink at 25 km. Therefore, an increase of water vapour would not necessarily lead to a large increase of nitric acid. Because the other reactive nitrogen gases NO , NO_2 , NO_3 and ClONO_2 are closely coupled with N_2O_5 reaction (1) may be a good candidate for the conversion of reactive nitrogen gases to nitric acid. The lifetime of N_2O_5 against reaction (1) is given by $t_1 = t_c/\gamma$ where γ is the reaction probability per collision. If reaction (1) was in fact the cause of the observed nitric acid increase and if we assume that reactive nitrogen other than HNO_3 was entirely in the form of N_2O_5 (implying rapid conversion of the photolysis product, NO_2 , to N_2O_5), we obtain the simplified steady-state equation

$$(\text{HNO}_3) = (\text{N}_2\text{O}_5) \frac{1}{t_1 J}$$

where *J* is the coefficient for HNO_3 photolysis, $\sim 1 \times 10^{-6} \text{ s}^{-1}$ at 25 km at middle latitudes in summer. A nearly complete conversion of reactive nitrogen to nitric acid, leading to, say, $(\text{HNO}_3)/(\text{N}_2\text{O}_5) = 10$, would require $t_1 = 10^5 \text{ s}$. For our upper limit of *A*, $2 \times 10^{-7} \text{ cm}^2 \text{ cm}^{-3}$, this implies $\gamma \geq 0.01$. Hence, our data set a lower limit on γ . As N_2O_5 represents only ~50% of the reactive nitrogen gases not in the form of HNO_3 , a more realistic lower limit is $\gamma \geq 0.02$.

Recent laboratory measurements¹¹⁻¹³ of γ for macroscopic $\text{H}_2\text{SO}_4/\text{H}_2\text{O}$ surfaces with a composition similar to that of stratospheric aerosols (~70% by mass H_2SO_4) range from 0.05 to 0.09. Our inferred upper limit of 0.02 is consistent with this. The laboratory measurements suggest that γ increases very sensitively with the H_2O fraction of H_2SO_4 - H_2O aerosols³. In turn, the H_2O fraction depends sensitively on the stratospheric water-vapour abundance and temperature, increasing for increasing water vapour and decreasing temperature. For a typical water-vapour abundance of 4 p.p.m.v. and a temperature of 225 K, as measured at 25 km on our balloon flight of 10 June 1982, the calculated H_2O mass fraction of H_2SO_4 - H_2O aerosols is ~24% similar to the laboratory measurements¹¹⁻¹³.

From the above, it seems likely that the observed increase in nitric acid concentration is strong evidence for the occurrence of aerosol-catalysed reactions in the stratospheric eruption cloud of El Chichón. If so, an important implication would be that nitric acid, probably formed by the reaction of N_2O_5 , does not remain in the aerosols, in contrast with recent speculation¹⁴. Passivation of reactive nitrogen by conversion to nitric acid may

lead to the NO_2 in the eruption cloud being <10% of the initial value. Because the eruption cloud encountered was quite thin, however, the total vertical NO_2 column would be reduced only by ~5%. Measurements of the total NO_2 column revealed decreases by ~15–20% after the El Chichón eruption in late 1982³. By that time, the eruption cloud was mostly thicker than the thin cloud encountered by our mass spectrometer. Owing to the decrease of NO_2 , the conversion of ClO to ClONO_2 and of BrO to BrONO_2 becomes less efficient and as a consequence, the abundances of ClO and BrO should increase, possibly leading to an increased ozone loss through the ClO - ClO and ClO - BrO catalytic cycles. The efficiency of the ClO - ClO cycle, however, is reduced at the mid-latitude stratospheric temperatures (around 25 km) owing to thermal dissociation of the ClO dimer. It is at least conceivable that other aerosol-catalysed reactions may have taken place on H_2SO_4 - H_2O aerosols in the eruption cloud, including reactions that activate ClONO_2 , HCl and BrONO_2 . If so, halogen activation would have been even more efficient, which may have led to even more ozone destruction. A future volcanic eruption with the strength of El Chichón during a period when the stratospheric abundances of chlorine and bromine are larger than in 1982 may cause substantially more ozone destruction. If there is no reduction in the chloro-fluorocarbon emissions, the total stratospheric chlorine will be doubled (compared to 1982) around the year 2000. The impact of a volcanic eruption cloud on ozone may then be quite severe, because the cloud has a typical lifetime of 1–2 years and may cover large parts of the Earth including the equatorial regions, where the unperturbed total ozone column is already relatively small. □

Received 18 April; accepted 6 September 1990.

1. Crutzen, P. & Arnold, F. *Nature* **324**, 651–655 (1986).
2. Toon, O. B., Hamill, P., Turco, R. P. & Pinto, J. *Geophys. Res. Lett.* **13**, 1284–1287 (1986).
3. Hofmann, D. & Solomon, S. *J. geophys. Res.* **94**, 5029–5041 (1989).
4. Jäger, H. & Wege, K. *J. atmos. Chem.* **10**, 273–287 (1990).
5. Adriani, A., Fiocco, G., Gobbi, G. P. & Congeduti, F. *J. geophys. Res.* **92**, 8365–8372 (1987).
6. Altorio, A. & Visconti, G. *Geophys. Res. Lett.* **10**, 27–30 (1983).
7. Arnold, F. & Bührke, Th. *Nature* **301**, 293–295 (1983).
8. Arnold, F. & Fabian, R. *Nature* **283**, 55–57 (1980).
9. Arnold, F., Fabian, R., Henschen, G. & Joos, W. *Planet. Space Sci.* **28**, 681–685 (1980).
10. Viggiano, A. A. & Arnold, F. *J. geophys. Res.* **88**, 1457–1462 (1983).
11. Mozurkewich, M. & Calvert, J. G. *J. geophys. Res.* **93**, 15889–15896 (1988).
12. Rossi, M., Malhotra, R. & Golden, D. M. *Geophys. Res. Lett.* **14**, 127–130 (1987).
13. Tolbert, M., Rossi, M. J. & Golden, D. M. *Geophys. Res. Lett.* **15**, 851–854 (1988).
14. Michelangeli, D. V., Allen, M. & Yung, Y. L. *J. geophys. Res.* **94**, 18429–18443 (1989).

ACKNOWLEDGEMENTS. We thank the staffs of CNES (Division Ballons, Aire-sur-L'Adour), KFA-Jülich and MPIK-Heidelberg. This project was funded by the Bundesministerium für Forschung und Technologie.

Intensification of recycling of organic matter at the sea floor near ocean margins

Richard A. Jahnke*, Clare E. Reimers† & Deborah B. Craven*

* Skidaway Institute of Oceanography, PO Box 13687, Savannah, Georgia 31416, USA

† Scripps Institution of Oceanography, A-020, La Jolla, California 92093, USA

QUANTIFICATION of the flux of organic carbon from the ocean surface to the deep ocean and sea floor is crucial to the prediction of future levels of atmospheric carbon dioxide and the interpretation of organic carbon variations in marine sediments. We report here the results of a study of benthic exchange and metabolism based on *in situ* benthic flux-chamber, *in situ* oxygen microelectrode and shipboard pore-water measurements performed off the

central California continental shelf. At the base of the continental slope, the rates of benthic carbon mineralization exceed $0.8 \text{ mol C m}^{-2} \text{ yr}^{-1}$. Seaward of the slope and rise, however, rates decrease to $<0.05 \text{ mol C m}^{-2} \text{ yr}^{-1}$ in the North Pacific central gyre. These results indicate that, across the northeast Pacific, about half of the input of organic carbon to the sea floor occurs within 500 km of the continental slope. Measured rates of benthic carbon oxidation near the continental margin exceed the fluxes of organic carbon determined from previous sediment-trap studies by a factor of about three. In this region, therefore, either traps significantly underestimate the mean flux of vertically sinking particulate organic carbon (perhaps by under-sampling important episodic events) or more than half of the organic-carbon input is due to processes that by-pass traps, such as near-bottom lateral or active biological transport.

In the following, we argue that vertical fluxes adjacent to continental margins have a critical role in ocean biogenic cycles. Intense vertical transport of organic matter in these regions may be supported by high rates of primary productivity¹, high ratios of new production to primary production^{2,3}, and possible lateral input from the adjacent continental shelf and slope^{4,5}. This view seemingly contradicts recent primary productivity and free-drifting sediment-trap studies that suggest that open-ocean regions dominate marine organic carbon cycling⁶.

To evaluate benthic rates of organic carbon oxidation on the continental slope and rise off central California (Fig. 1), *in situ* benthic chamber incubations were performed at eight sites using the Benthic Experimental Chamber Instrument (BECI)⁷ in November 1987, July 1988 and June 1989. Oxygen (Fig. 2a) and NO_3^- (Fig. 2b) concentrations were measured to precisions of $\pm 1\%$ and $\pm 2\%$, respectively⁸. During the same period, *in situ* interfacial profiles of dissolved O_2 (Fig. 2c) and resistivity were measured at 13 sites using the *in situ* microprofiler (IMP)⁹. Pore-water samples were extracted on board ship from box cores (BC) using a pressurized-core barrel technique¹⁰ and analysed for SO_4^{2-} ($\pm 1.5\%$; Fig. 2d) by the method of Tabatabai¹¹.

Benthic O_2 fluxes estimated from chamber time series and microelectrode results are displayed in Fig. 3a along with an O_2 hydrographic profile measured at GEOSECS 202, which illustrates the well developed oxygen-minimum zone in this region. There is a large range in flux estimate even between measurements made less than a kilometre apart at similar water depths. This variability was expected considering the uncertainties of both techniques, the heterogeneous nature of surface sediments, and the patchiness of macrobenthic organisms in this region^{12,13}. We note, however, that the intrasite measurement variability is smaller than the overall range of values and that there is no consistent difference between the techniques for estimating the benthic O_2 flux.

In general, the lowest O_2 fluxes are observed in the O_2 -minimum zone. Sea-floor O_2 fluxes tend to increase with increasing water depth and increasing bottom-water O_2 to the upper continental rise. Farther offshore, at depths of $\sim 4,400 \text{ m}$, lower flux values are observed, but these values are still larger than those measured in the O_2 -minimum.

The general correlation observed between bottom-water O_2 and benthic O_2 flux over most of the depths sampled could result because solute fluxes are limited by diffusive transport across the sediment/water interface or because organic carbon fluxes are higher seaward of the O_2 -minimum zone. Estimates of the diffusive sublayer thickness made by alabaster dissolution¹⁴ and O_2 microelectrode^{15,16} at a subset of these stations are $\leq 1 \text{ mm}$. Oxygen gradients measured in this data set extend to $>4 \text{ mm}$ depth (Fig. 2c). Thus, calculations indicate that the rates of O_2 -consuming reactions in the sediments limit the benthic O_2 flux, not transport across the sublayer. The general agreement between calculated and measured fluxes support this conclusion. If diffusive transport were limiting, the chamber fluxes would be systematically biased by the chamber stirring

rate, which was set nominally to maintain a diffusive sublayer thickness of $200\text{--}400 \mu\text{m}$ in the chamber¹⁷.

Benthic NO_3^- fluxes tend to be near zero except in the O_2 -minimum zone (Fig. 3b). If we assume that at all sites studied, the decomposing organic matter has a Redfield C:N atomic ratio of 6.6, and that the organic nitrogen released in oxic surface sediments is oxidized to NO_3^- , we would expect a flux of NO_3^- from the sediments of $0.5\text{--}0.8 \text{ mol m}^{-2} \text{ yr}^{-1}$. Because the results are consistently near zero or slightly negative, either there is significant denitrification in these sediments or dissolved organic nitrogen is diffusing out of the sediments at a rate equal to the particulate nitrogen input. In the following calculations, we assume that denitrification accounts for the missing NO_3^- flux. Benthic SO_4^{2-} fluxes estimated from the pore-water gradients are $<0.1 \text{ mol m}^{-2} \text{ yr}^{-1}$ at all sites.

Primary productivities reported by the VERTEX⁶ program and those measured in the central California margin area by the California Cooperative Oceanic Fisheries Investigations (CalCOFI)¹⁸ program for 1984 (the same year that the VERTEX

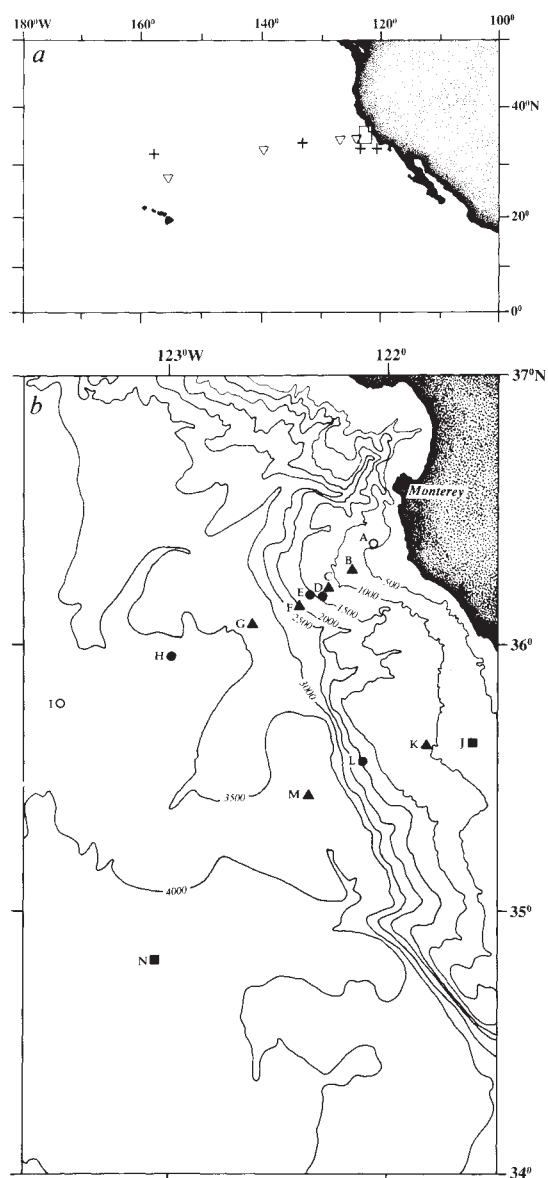
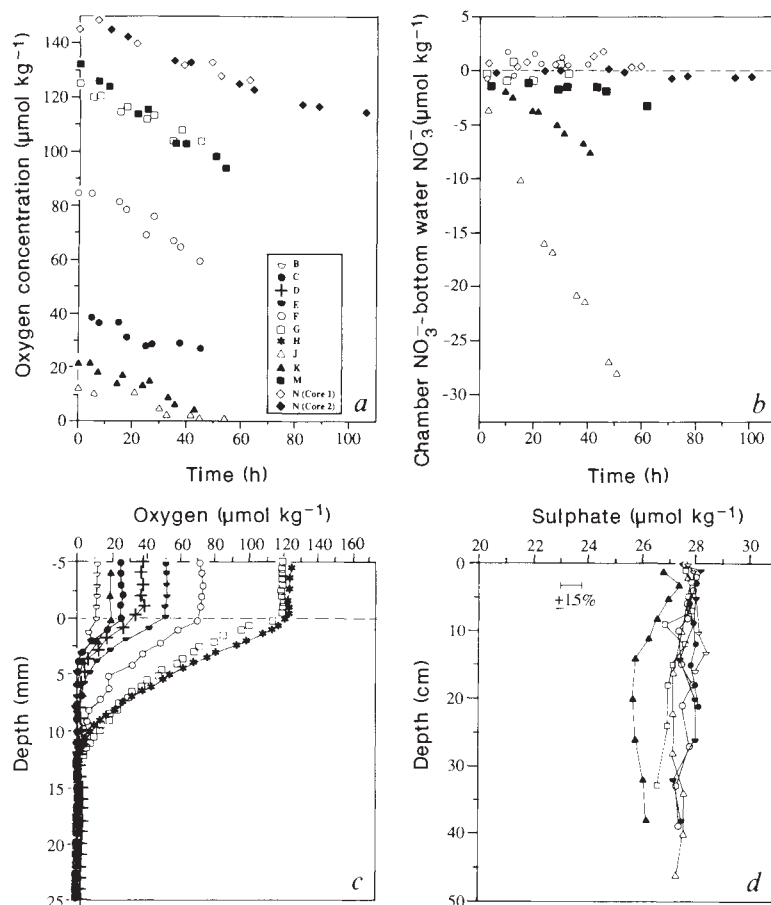


FIG. 1 a, Locations of previous benthic chamber¹⁸ and ∇ , pore water²⁸ results for the eastern North Pacific. The box indicates the region shown in b. b, Study sites on the central Californian continental margin. Symbols indicate sampling techniques used at a given site: $\blacktriangle$, BECI, IMP and BC; $\bullet$, IMP and BC; $\blacksquare$, BECI and BC.

FIG. 2 *a*, Oxygen and *b*, nitrate concentration as a function of time during chamber incubations. Individual stations for each panel are represented by the symbols indicated in *a*. For clarity, NO_3^- results are presented as the difference between the measured chamber and bottom-water concentrations. Benthic fluxes were estimated from the slope of the concentration-time results (determined by linear regression of the data) divided by the chamber volume and normalized to a unit area. *c*, Representative oxygen profiles measured by microelectrode at a subset of the study sites. Up to 6 profiles were measured at each station. Oxygen fluxes were calculated for each profile and then a station average determined. The calculations assume one-dimensional diffusive transport into the sediment as described by Fick's first law

$$J = \phi \frac{D}{\theta} \frac{\delta O_2}{\delta z} \bigg|_{z=0} \approx \frac{D}{F} \frac{\delta O_2}{\delta z} \bigg|_{\text{max}}$$

where J is the benthic flux, D is the molecular diffusion coefficient of O_2 in sea water at bottom-water temperature, ϕ is the sediment porosity, θ is the sediment tortuosity, F is the formation factor [electrical resistivity of bottom sediment]/[resistivity of bottom water] and $\delta O_2/\delta z|_{\text{max}}$ is the steepest gradient measured in the upper 1–3 mm of the oxygen profiles. *d*, Sulphate profiles: fluxes were also calculated from Fick's first law using $\phi^2 D$ as an estimate of the effective diffusion coefficient where ϕ is the porosity and D is the molecular diffusion coefficient in free solution.



I site was studied) are displayed in Fig. 4*a*. The solid line represents a seasonally weighted mean of the ocean margin data and the dashed line is the open-ocean composite mean reported by the VERTEX program⁶. In Fig. 4*b*, we indicate the organic carbon flux to the sea floor estimated from the regressions reported for the VERTEX sediment-trap collections and the mean surface-water productivities displayed in Fig. 4*a*.

Organic carbon oxidation rates at the sea floor were calculated using the estimated fluxes of oxidants (O_2 , NO_3^- and SO_4^{2-}) into the sediments and assuming Redfield stoichiometry. At sites where only pore-water profiles of O_2 and SO_4^{2-} were measured, NO_3^- fluxes were estimated by extrapolating the chamber measurements (Fig. 3*b*) to the appropriate water depth. The water depth at each sampling site is also displayed (Fig. 4*c*) so that the relationship between each location and the continental slope is clear.

In general, the highest benthic mineralization rates and inferred organic-carbon input rates occur from the base of the continental slope seaward for a distance of 100 km. These rates, like the O_2 fluxes, are higher than those measured at shallower depths where the O_2 -minimum impinges on the sediments. Farther offshore, benthic mineralization rates also decrease.

Sea-floor mineralization rates exceed the input rates estimated from sediment traps by about a factor of three, over an area extending to ~300 km from the continental slope. We do not attribute this difference to the relationship used to extrapolate the VERTEX results from the deepest measurement depth (2,000 m) to the sea floor, as benthic fluxes exceed the sediment-trap value at 2,000 m at the VERTEX I site by a factor of two. Similar discrepancies have also been observed adjacent to the California Borderland area¹⁹ and in the Panama basin²⁰. (A calculation error in ref. 20 accounts for the erroneous agreement between benthic flux and the particulate input reported there.) The difference between benthic and sediment-trap fluxes indi-

cated in Fig. 4*b* is probably a minimum estimate because burial has not been considered.

Several possibilities may account for the higher sea-floor-derived carbon fluxes. VERTEX results could represent an anomalously low-flux period and not reflect mean annual rates. The surface-water primary production rates measured during the VERTEX trap deployments agree well with the CalCOFI summer results and were slightly higher than the annual mean value, however, suggesting that if anything, one might expect measured fluxes to be higher than the annual average, not lower. Alternatively, episodic sedimentation events not captured by

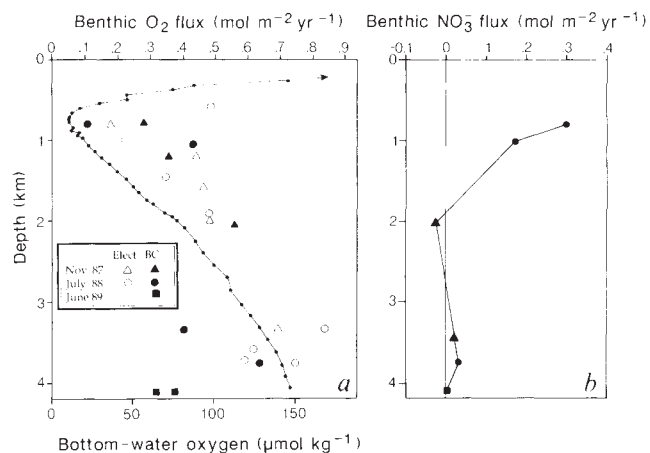


FIG. 3 *a*, Benthic oxygen and *b*, nitrate fluxes as a function of water depth for the indicated expeditions and techniques. Fluxes into the sediments are represented by positive values. Included in *a* is the GEOSECS 202 oxygen hydrographic profile (line).

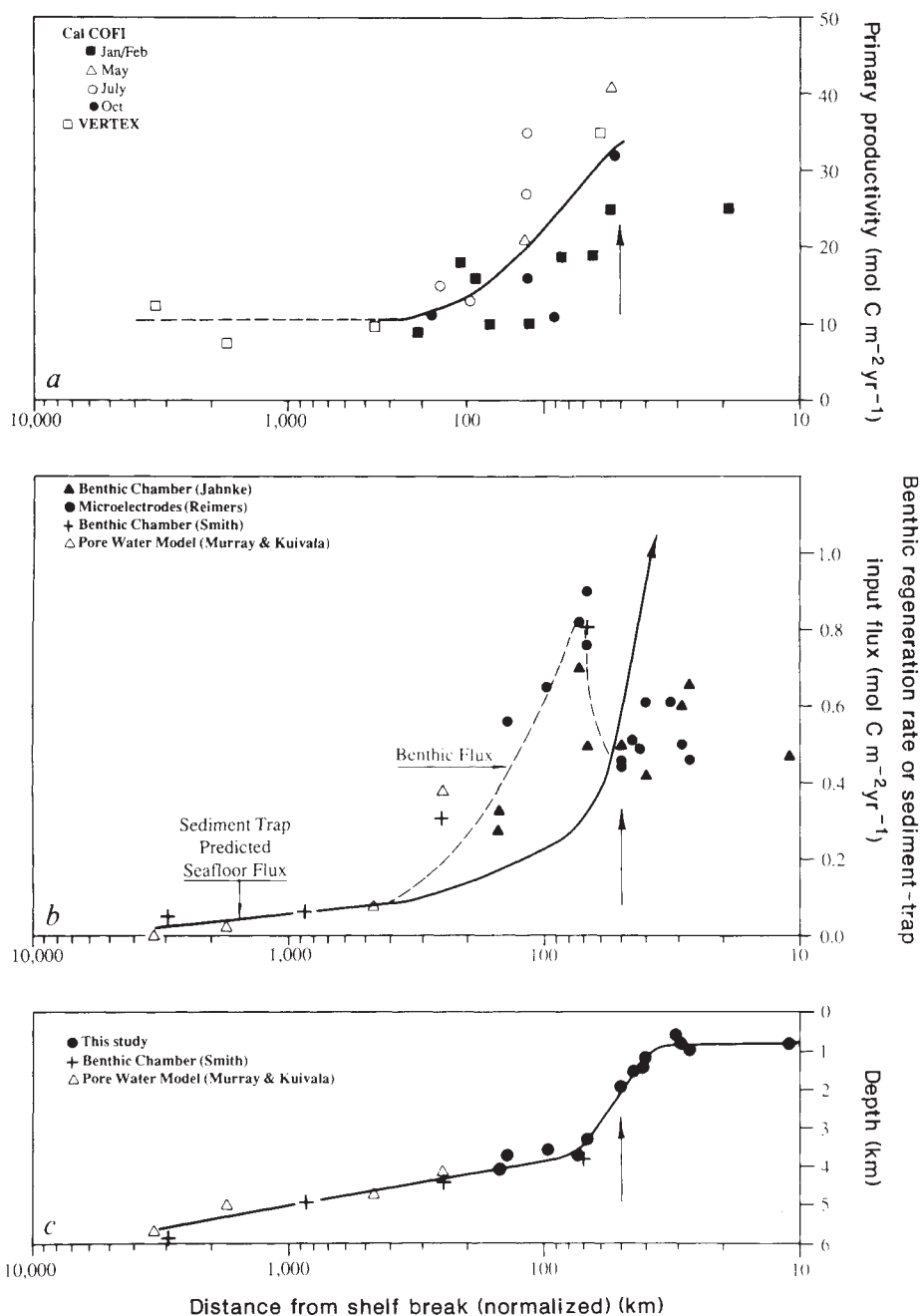


FIG. 4 *a*, Surface-water primary productivity for the indicated samplings plotted against normalized seaward distance from the shelf break. To account for differences in margin morphology, each sampling location has been adjusted such that the centre of the continental slope, assumed to be represented by the 2,000-m isobath (the arrow in each panel), is 50 km from the continental shelf. *b*, Estimated total sea-floor organic carbon regeneration rate as a function of distance from shelf. Solid line represents the particulate organic carbon flux to the sea floor estimated from regressions of the VERTEX sediment trap results⁶ at the indicated sites applied to the surface-water productivity at each location displayed in *a*. *c*, Water depth of each sampling location.

month-long sediment-trap deployments may comprise a significant portion of the mean annual flux.

Another possibility is that the passive settling of particulate matter captured by sediment traps is not the dominant mechanism by which organic material is transported to the deep-ocean near margins. Active migration of benthic organisms has been suggested to represent an important pathway by which metabolic energy is transported down the water column to the benthos^{21,22}. In addition, near-bottom lateral transport may contribute marine and terrestrial organic materials to the continental-margin sea floor. Lateral input has been well documented for lithogenic particles²³⁻²⁶. Recent studies have documented the occurrence of down-slope transport of organic matter^{4,27}, but the quantitative significance of such processes has not been established. Although not yet fully explained, the difference between trap flux and benthic flux observed near margins may provide a useful means of defining ocean boundary regions.

The total flux to the sea floor, estimated by integrating under the curve derived from the sediment-trap results (Fig. 4*b*)

accounts for only 48% of the total benthic flux across the transect. Estimates of organic carbon cycling based solely on sediment-trap collections may therefore significantly underestimate basin-wide deep-ocean input rates. Although the reported surface-water primary productivity varies only by a factor of 3-4 between the open ocean and continental-margin regions⁶, benthic fluxes vary by more than a factor of 20. Because open-ocean area is much larger than margin area, models of ocean fluxes based on surface-water primary production and relationships between sediment-trap fluxes and depth suggest that open-ocean fluxes dominate carbon cycling. Benthic fluxes, however, reveal that ~53% of the total measured sea-floor flux across this >3,000 km transect occurs within 500 km of the continental slope. □

Received 16 July; accepted 5 September 1990.

- Berger, W. H., Fischer, K., Lai, C. & Wu, G. *Undersea Res. Symp.*, 88-1 (ed. Agegian, C.) 131-176 (National Oceanic and Atmospheric Administration, Silver Springs, 1988).
- Eppley, R. W. & Peterson, B. J. *Nature* **282**, 677-680 (1979).
- Harrison, W. G., Platt, T. & Lewis, M. R. *J. Plank. Res.* **9**, 235-248 (1987).

4. Yoder, J. A. & Ishimaru, T. *Cont. Shelf. Res.* **9**, 547–553 (1989).
5. Jahnke, R. A. & Jackson, G. A. *Nature* **329**, 621–623 (1987).
6. Martin, J. H., Knauer, G. A., Karl, D. M. & Broenkow, W. W. *Deep Sea Res.* **34**, 267–285 (1987).
7. Jahnke, R. A. & Christiansen, M. B. *Deep Sea Res.* **36**, 625–637 (1989).
8. Strickland, J. D. H. & Parsons, T. R. *Bull.* 167 (Fish Res. Board Can., Ottawa, 1972).
9. Reimers, C. E. *Deep Sea Res.* **34**, 2019–2035 (1987).
10. Jahnke, R. A. *Limnol. Oceanogr.* **33**, 483–487 (1988).
11. Tabatabai, M. A. *Envir. Lett.* **7**, 237–243 (1974).
12. Thompson, J. B., Mullins, H. T., Newton, C. R. & Vercoutere, T. L. *Lethaia* **18**, 167–179 (1985).
13. Vercoutere, T. L., Mullins, H. T., McDougall, K. & Thompson, J. B. *J. sedim. Petrol.* **57**, 709–722 (1987).
14. Anderson, R., Fleisher, M. & Santschi, P. H. (abstr.) *Trans. Am. geophys. Un.* **71**, 124 (1990).
15. Reimers, C. E. (abstr.) *Trans. Am. geophys. Un.* **69**, 1123–1124 (1988).
16. Archer, D., Emerson, S. & Smith, C. R. *Nature* **340**, 623–626 (1989).
17. Buchholtz-ten Brink, M. R., Gust, G. & Chavis, D. *Deep Sea Res.* **36**, 1083–1101 (1989).
18. California Cooperative Oceanic Fisheries Investigations *Data Rep.* *SIO Ref.* **84-18**, 84-23, 84-25, 84-30 and 85-1 (Scripps Institution of Oceanography, La Jolla).
19. Smith, K. L. Jr *Limnol. Oceanogr.* **32**, 201–220 (1987).
20. Cole, J. J., Honjo, S. & Erez, J. *Nature* **327**, 703–704 (1987).
21. Smith, K. L. Jr, Alexandrou, D. & Edelman, J. L. *Deep Sea Res.* **36**, 1427–1441 (1989).
22. Houston, K. A. & Haedrich, R. L. *Mar. Biol.* **92**, 563–574 (1986).
23. Honjo, S., Spencer, D. W. & Farrington, J. W. *Science* **216**, 516–518 (1982).
24. Richardson, M. J. *Deep Sea Res.* **34**, 1304–1329 (1987).
25. Thorpe, S. A. & White, M. *Deep Sea Res.* **35**, 1665–1671 (1988).
26. Richardson, M. J. & Hollister, C. D. *J. mar. Res.* **45**, 175–200 (1987).
27. Grant, J., Volckaert, F., Roberts-Regan, D. L. *Cont. Shelf Res.* **7**, 1123–1138 (1987).
28. Murray, J. W. & Kuivila, K. M. *Deep Sea Res.* **37**, 59–80 (1990).

ACKNOWLEDGEMENTS. We thank the captain and crew of the R/V *New Horizon* and M. Christiansen for assistance in the field. K. L. Smith Jr provided ship time in June 1989. Graphics were finalized by A. Boyette and L. Land prepared the final drafts of the manuscript. Ship time in 1987 and 1988 and other support were provided by the NSF.

Aerobic respiration in the Archaean?

Kenneth M. Towe

Department of Paleobiology, Smithsonian Institution,
Washington, DC 20560, USA

THE Earth's atmosphere during the Archaean era (3,800–2,500 Myr ago) is generally thought to have been anoxic, with the partial pressure of atmospheric oxygen about 10^{-12} times the present value¹. In the absence of aerobic consumption of oxygen produced by photosynthesis in the ocean, the major sink for this oxygen would have been oxidation of dissolved Fe(II). Atmospheric oxygen would also be removed by the oxidation of biogenic methane. But even very low estimates of global primary productivity, obtained from the amounts of organic carbon preserved in Archaean rocks, seem to require the sedimentation of an unrealistically large amount of iron and the oxidation of too much methane if global anoxia was to be maintained. I therefore suggest that aerobic respiration must have developed early in the Archaean to prevent a build-up of atmospheric oxygen before the Proterozoic. An atmosphere that contained a low (0.2–0.4%) but stable proportion of oxygen is required.

In a classical Archaean environment the oxygen produced by net photosynthesis is not used by the aerobic heterotrophic processes (aerobiosis) that today rapidly recycle it into carbon dioxide, nor is methane or ammonia oxidized biologically in the water column because methylophony and bacterial nitrification both require molecular oxygen. Without aerobic respiration, which rapidly remineralizes ~99% of the organic carbon produced by net marine photosynthesis today, no oxygen-minimum zone due to carbon oxidation would exist in the Archaean ocean, nor would sediment/water interfaces be areas of competition and transition for aerobic and anaerobic processes. Organic carbon would have been subjected to anaerobic remineralization, which today accounts for <1% of the carbon recycled in the oceans², or would have been remaindered into the sediment record. The anaerobic cycles during the Archaean would themselves have been different from those of today because, with limited nitrate³ as a secondary oxidant, neither the process of bacterial denitrification nor nitrate ammonification would have been significant. Bacterial iron and sulphate reduction occurred, but in this anoxic world, organic carbon

was recycled primarily by the processes of fermentation and carbonate-CO₂ reduction leading to methanogenesis. The oxygen produced by early cyanobacterial photosynthesis (and not used by gross respiration) would be released into the surrounding water to meet ferrous iron and be buried as ferric oxide minerals. Little oxygen was left to diffuse from the hydrosphere to the atmosphere⁴. Therefore, without aerobic heterotrophic recycling of organic carbon, the burial of iron (and sulphur), not the burial of organic matter, would be the dominant control over the oxygen transferred to the atmosphere.

No one knows how much carbon was produced annually by photosynthesis in the Archaean oceans, nor do we know how much of the organic carbon was recycled or how much was buried with sediments. Neither do we know very accurately how much iron was oxidized and deposited each year in the sediment record nor what percentage of the carbon was converted anaerobically into methane and what percentage into carbon dioxide. Nevertheless, quantitative, if speculative approaches to such problems have proven useful^{1,3,5–9}.

The average total organic carbon content of Archaean sedimentary rocks is more or less in line with that found in younger rocks^{1,7,10}. Regardless of whether or not the fraction of organic carbon buried has changed significantly over most of geological time, a value of about 0.5–0.6% carbon by weight is considered to be as reasonable for the average Archaean sedimentary rock as it is for many Phanerozoic sediments¹⁰. Accepting an estimated global rate of sedimentation and sediment recycling of $\sim 10^{16}$ g yr⁻¹ (ref. 11), an average $5\text{--}6 \times 10^{13}$ g ($4\text{--}5 \times 10^{12}$ mol) of organic carbon should have been buried each year (after losses from post-depositional diagenesis). Other estimates for the Archaean have been double this amount (10^{13} mol yr⁻¹; ref. 1), and therefore closer to the modern burial rate. The carbon preserved and buried annually during the Archaean, and now present as kerogen, represented some fraction of the carbon produced each year by net photosynthesis, but the percentages of recycling and preservation in the Archaean are unknown and must be assumed. Today, the overall rate of preservation in the world oceans is only 0.3% of the productivity². In some offshore open marine areas it may be higher, perhaps averaging 1%, whereas in most nearshore settings and in the high-productivity zones of upwelling the preservation percentages can be as high as 10–20%¹². In the euxinic Black Sea, after correction for the influence of terrigenous carbon, the rate of preservation is 4–5%. Palaeoproductivity estimates for the mid-Cretaceous open oceans assumed a preservation of 2% in the deeper anoxic regions¹³. It is probable that overall Archaean recycling was not as efficient as it is today, nor perhaps even as efficient as suggested for the mid-Cretaceous anoxic areas. Much more of the carbon produced may have been buried in the Archaean, perhaps as much as the 10–20% that is now buried in today's strongly anoxic locales¹². If I assume that the carbon buried during the Archaean averaged 20% of the carbon produced and that $5\text{--}6 \times 10^{13}$ g of carbon were buried each year in sediments, the entire global primary production would have been in the range of $25\text{--}30 \times 10^{13}$ g yr⁻¹. If aerobic processes do not occur, the 80% of this productivity not buried ($20\text{--}24 \times 10^{13}$ g) must have been recycled anaerobically.

Depending on the areal extent of the Archaean oceans and the degree of development of the continents¹¹, this estimated average global productivity represents a meagre $0.5\text{--}0.7$ g C m⁻² yr⁻¹. By modern standards this is exceptionally low, <1% of global marine net productivity of 140 g C m⁻² yr⁻¹ (ref. 14). The total estimated global Archaean productivity is the equivalent of $2.1\text{--}2.5 \times 10^{13}$ mol of organic carbon produced annually by photosynthesis. Because the net reaction for oxygenic photosynthesis is $\text{H}_2\text{O} + \text{CO}_2 \rightarrow \text{CH}_2\text{O} + \text{O}_2$, a similar number of moles of oxygen should have been released into the Archaean ocean each year. Without the benefit of heterotrophic carbon oxidation by respiration, the removal of each mole of

molecular oxygen would require four moles of ferrous iron: $4\text{Fe}^{2+} + \text{O}_2 + 4\text{H}_2\text{O} \rightarrow 2\text{Fe}_2\text{O}_3 + 8\text{H}^+$. If iron had been the sole sink for the photosynthetically produced oxygen, $8.4\text{--}10 \times 10^{13}$ mol of total iron would have been oxidized annually, and therefore $470\text{--}560 \times 10^{13}$ g of iron would have been deposited with sediments each year. Dividing by the global rate of sedimentation and recycling reveals that the average Archaean sediment should contain a staggering 47–56% total iron (67–80% Fe_2O_3). The global average iron concentration for all sediments is only ~5% total iron⁷, whereas the richer Precambrian banded iron formations contain an average 30% total iron. Clearly, without aerobic respiratory recycling an enormous amount of iron is needed to offset even a very low cyanobacterial productivity, and it is not present even in the richest deposits. The average amount of organic carbon preserved in Archaean rocks is derived from a widely accepted, if uncertain, value¹⁰ so that altering the global sedimentation rate will not affect the percentage of iron obtained in this manner. The organic carbon preserved in the Archaean rock record therefore suggests that if the Earth was indeed anoxic then iron was not the main oxygen sink. The productivity (as a result of the arbitrarily high average preservation rate chosen) is already very low and at odds with classical views of the early Earth as a prolific microbial ecosystem limited only by nutrient availability¹⁰. This 'small' production and oxygen source cannot be further minimized by arguing that most of the global primary productivity was the result of bacterial anoxygenic photosynthesis¹⁵ because the low solubility of iron sulphide requires that, in an ocean replete with dissolved ferrous iron, the concentration of sulphide should have been negligible ($10^{-6}\text{--}10^{-8}$ M)⁸. These equilibrium values are several orders of magnitude too low to support either the needs of the photosynthetic bacteria or to initiate the process facultatively in cyanobacteria¹⁶.

Using the same sedimentation rate and accepting 5% total iron as a reasonable estimate for the global average iron concentration in all sediments, I obtain 5×10^{14} g Fe yr^{-1} . The oxygen consumed by this total iron would be 2.2×10^{12} mol O_2 yr^{-1} . But with $21\text{--}25 \times 10^{12}$ mol of photosynthetic oxygen being produced annually, >90% of it remains to be accounted for if the atmosphere and hydrosphere are to remain anoxic. Little, if any, loss owing to the weathering of 'old' kerogen exposed by erosion is likely and, after its deep-ocean titration^{3,8}, the amount of sulphide remaining in equilibrium with the dissolved ferrous iron is also too small to be of concern. Therefore, with no important terrestrial 'sinks' remaining accessible to consume the remaining $19.3\text{--}22.8 \times 10^{12}$ mol of oxygen the burden must be passed from the hydrosphere and lithosphere to components of a reducing atmosphere or the Earth would soon become oxidizing^{1,8}.

In the atmosphere a potentially important sink is the biogenic methane produced as a result of the anaerobic recycling of organic carbon. The net reaction for methanogenesis is $2\text{CH}_2\text{O} \rightarrow \text{CH}_4 + \text{CO}_2$. In today's marine environment, methanogenesis represents only 0.1% of primary productivity² and little of the methane escapes to the atmosphere. With 80% of the Archaean productivity recycled anaerobically each year (and provided that bioconversion to methane was complete) the entire $1.7\text{--}2.0 \times 10^{13}$ mol of recycled organic carbon could yield at most $8.5\text{--}10 \times 10^{12}$ mol of methane. (Previous quantitative estimates of the Archaean microbial methane flux assumed a much larger annual productivity and are 30 times greater⁹.) Methane could consume twice this number of moles, or $17\text{--}20 \times 10^{12}$ mol, of molecular oxygen, provided that all the methane produced was oxidized to CO_2 . Even with the remote possibility that both complete bioconversion and complete oxidation took place, this would still leave 2.8×10^{12} mol of oxygen to be removed from the atmosphere. The last remaining sink would be the reduced gases added each year to the atmosphere by hydrothermal and volcanic activity. Made up primarily of hydrogen, carbon monoxide and reduced sulphur gases, this sink has been estimated to

represent an Archaean demand for $\sim 10^{12}$ mol O_2 yr^{-1} (ref. 8). A threefold increase in the estimated hydrothermal and volcanic activity would thus be required to complete the maintenance of a balanced anoxic condition.

Without aerobic processes to dominate the remineralization of organic carbon a substantial flux of methane to the Archaean atmosphere was both necessary and unavoidable if carbon was to be recycled and if the Earth was to remain anoxic. Extensive global methane production is supported, at first glance, by the fact that the archaeobacterial methanogens are phylogenetically primitive organisms that flourish under conditions of strong anoxia and low dissolved sulphate. On the other hand, there are important obstacles in the path of global methanogenesis being the predominant organic-carbon recycling mechanism in an anoxic world. Notwithstanding an evolutionary antiquity for the methanogens, their early appearance does not automatically provide for the recycling of a large percentage of the global productivity. On the contrary, all methanogens are ecologically restricted in their ability to deal readily with primary organic materials^{2,17,18}. Their choice of substrates is limited to acetate, hydrogen- CO_2 and methylamines¹⁹. This restricts the overall recycling of primary organic carbon, and hence the biogenic transformation of organic matter into methane and then into carbon dioxide. Methanogens are therefore dependent on a close association with a variety of other organisms that can break down the more complex organic molecules into forms the methanogens can use. These eubacteria coexist with the methanogens in a symbiotic relationship in which a complex process known as interspecies hydrogen transfer takes place¹⁷. It is possible to visualize this complex ecosystem operating in ancient microbial mat communities or in local, sedimentologically quiet locales. It is difficult, however, to accept that it could have been the dominant process taking place in the dynamic and unstable world of the average open marine or nearshore environment, where the bulk of the organic carbon is produced and recycled and where most sediments are deposited¹². Furthermore, biogenic methane is unusually depleted in the heavy isotope of carbon ($\delta^{13}\text{C} = -70$ to -80% , ref. 20; $\delta^{13}\text{C} = [({}^{13}\text{C}/{}^{12}\text{C})_{\text{sample}}/({}^{13}\text{C}/{}^{12}\text{C})_{\text{standard}}] - 1$ where the standard is PDB). Residual total carbon dioxide in the local environment is subject to enrichment in the heavy isotope as light methane is produced and escapes to the atmosphere. Because methane production tends to raise the pH, it is reasonable to assume that if it had been an important recycling pathway in the Archaean, more than a few locally precipitated, isotopically heavy carbonates would be expected, especially in closed or partially closed systems¹⁸. Such anomalously heavy carbonates have not been reported from the Archaean^{10,21}.

A plausible and simple solution to the organic-carbon/iron/methane dilemma is to abandon the idea of an anoxic Archaean atmosphere and replace it with one of a low, but stable oxygen atmosphere and a stratified ocean with oxidized surface waters overlying deep-water sources of ferrous iron^{7,22}. Most importantly, if the Archaean atmosphere contained oxygen at nearly 1–2% of the present atmospheric level (ref. 7), such conditions would be sufficient to offer at least the minimum level of oxygen²³ (0.2–0.4%) necessary to support aerobic heterotrophic recycling of organic carbon and thereby remove a large percentage of the photosynthetic oxygen released each year. In addition, the nitrogen cycle then becomes viable, as does the process of methylotrophy, a mechanism for oxidizing methane and generating the isotopically light organic carbon observed in a number of late Archaean sediments²⁴. In this global Archaean ecosystem, the preservation and burial of organic carbon would have its accepted modern role of controlling the bulk of the oxygen 'leaking' to the atmosphere²⁵. An oceanic iron reservoir and volcanic gases would remain important, if subordinate brakes on oxygen build-up in the atmosphere. Locally shallow upwelling (coastal, wind driven) could bring nutrients to support higher productivities nearshore

and associated offshore downwelling of surface waters could bring oxygen to meet deep-water stratified ferrous iron for the deposition of banded iron formations in distal areas near their hydrothermal sources and away from clastic sediment input (manuscript in preparation).

The early Archaean evolution of oxygenic photosynthesis²⁶ means that, as incongruous or unexpected as it may seem, the early evolution of aerobic respiration and heterotrophy in fully aerobic environments is necessary to help keep atmospheric oxygen levels low, terrestrial red beds few in number, detrital uraninites from being easily oxidized, the carbon isotope picture in reasonable balance, and the Precambrian deep oceans from being swept free of iron for hundreds of millions of years until the Proterozoic. It has already been suggested that the evolution of aerobic respiration could have either preceded the evolution of oxygenic photosynthesis²⁷ or have been closely contemporaneous with it²⁶, and there are biochemical²⁸ arguments to support the presence of low levels of oxygen early on. In this respect, the negative cerium anomalies reported in the rare-earth-element analyses of rocks from the banded iron formation of the 3,800-Myr Isua Formation in Greenland²⁹ are noteworthy, especially if the development of such anomalies required microbial oxidation of Ce(III) to insoluble Ce(IV), as they appear to do today³⁰. If there was indeed a shift to significantly higher oxygen levels during the Proterozoic^{1,7,31}, this shift may have been the result of tectonic changes that lowered volcanic activity and the input of iron and reduced gases into the oceans, dramatic changes in photosynthetic productivity³², or a change in the rate of preservation of organic carbon. As declining hydrothermal activity was drawing to a close the era during which most banded iron formations were deposited, the oxygen added to the atmosphere through carbon burial was no longer absorbed by as many volcanic gases. This would accelerate the development of a more effective ozone screen, leading to a broadening of the planktonic realm and eventually to the origin of metaphytes and metazoans. □

Received 19 April; accepted 4 September 1990.

- Kasting, J. F. *Precamb. Res.* **34**, 205–229 (1987).
- Henrichs, S. M. & Reeburgh, W. S. *Geomicrobiol. J.* **5**, 191–237 (1987).
- Kasting, J. F. & Walker, J. C. G. *J. geophys. Res.* **86**, 1147–1158 (1981).
- Cloud, P. E. *Econ. Geol.* **68**, 1135–1143 (1973).
- Garrels, R. M. & Perry, E. A. Jr in *The Sea*, Vol. 5 (ed. Goldberg, E. D.) 303–336 (Wiley, New York, 1974).
- Walker, J. C. G. in *Handbook of Environmental Chemistry* Vol. 1, Part A (ed. Hutzinger, O.) 87–104 (Springer, Berlin, 1980).
- Holland, H. D. *The Chemical Evolution of the Atmospheres and Oceans* (Princeton University Press, 1984).
- Walker, J. C. G. & Brimblecombe, P. *Precamb. Res.* **28**, 205–222 (1985).
- Kasting, J. F., Zahnle, K. J. & Walker, J. C. G. *Precamb. Res.* **20**, 121–148 (1983).
- Schidlowski, M. *Nature* **333**, 313–318 (1988).
- Veizer, J. & Jansen, S. L. *J. Geol.* **93**, 625–643 (1985).
- Müller, P. J. & Suess, E. *Deep Sea Res.* **A26**, 1347–1362 (1979).
- Bralower, T. J. & Thierstein, H. R. *Geology* **12**, 614–618 (1984).
- Martin, J. H., Knauer, G. A., Karl, D. M. & Broenkow, W. W. *Deep Sea Res.* **34**, 267–285 (1987).
- Reimer, T. O., Barghoorn, E. S. & Margulies, L. *Precamb. Res.* **9**, 93–104 (1979).
- Towe, K. M. *Origins of Life* **15**, 235–250 (1985).
- Wolin, M. J. & Miller, T. L. *Geomicrobiol. J.* **5**, 239–259 (1987).
- Claypool, G. E. & Kaplan, I. R. in *Natural Gases in Marine Sediments* (ed. Kaplan, I. R.) 99–139 (Plenum, New York, 1974).
- Winfrey, M. R. & Ward, D. M. *Appl. env. Microbiol.* **45**, 193–199 (1983).
- Rosenfeld, W. D. & Silverman, S. R. *Science* **130**, 1658–1659 (1959).
- Schidlowski, M., Hayes, J. M. & Kaplan, I. R. in *Earth's Earliest Biosphere* (ed. Schopf, J. W.) 149–186 (Princeton University Press, 1983).
- Drever, J. I. *Geol. Soc. Am. Bull.* **85**, 1099–1106 (1974).
- Chapman, D. J. & Schopf, J. W. in *Earth's Earliest Biosphere* (ed. Schopf, J. W.) 302–320 (Princeton University Press, 1983).
- Hayes, J. M. in *Earth's Earliest Biosphere* (ed. Schopf, J. W.) 291–301 (Princeton University Press, 1983).
- Berner, R. A. & Canfield, D. E. *Am. J. Sci.* **289**, 333–361 (1989).
- Schopf, J. W. in *Molecular Evolution and the Fossil Record* (ed. Broadhead, T. W.) 89–97 (University of Tennessee, Knoxville, 1988).
- Schwartz, R. M. & Dayhoff, M. O. *Science* **199**, 395–403 (1978).
- Towe, K. M. in *Molecular Evolution and the Fossil Record* (ed. Broadhead, T. W.) 114–129 (University of Tennessee, Knoxville, 1988).
- Dymek, R. F. & Klein, C. *Precamb. Res.* **39**, 247–302 (1988).
- Moffett, J. W. *Nature* **345**, 421–423 (1990).
- Walker, J. C. G., Klein, C., Schidlowski, M., Stevenson, D. J. & Walter, M. W. in *Earth's Earliest Biosphere* (ed. Schopf, J. W.) 260–290 (Princeton University Press, 1983).
- Knoll, A. H. *Origins of Life* **9**, 313–327 (1979).

ACKNOWLEDGEMENTS. I thank R. A. Berner for suggestions and a review, and my colleagues in the Precambrian Paleobiology Research Group and J. W. Schopf, B. Runnegar and R. S. Hoffman for discussions.

Some consequences of a proposed fractal nature of continental faulting

Ph. Davy*, A. Sornette† & D. Sornette†

* Centre Armorica d'Etude Structurale des Socles, Université de Rennes 1, 35042 Rennes Cedex, France

† Laboratoire de Physique de la Matière Condensée, CNRS URA 190, Faculté des Sciences, Parc Valrose, 06034 Nice Cedex, France

A KEY question in the study of continental deformation concerns the preservation of undeformed regions, sometimes quite large, in the midst of otherwise penetrative deformation. Some authors¹ have interpreted such regions as being relatively strong, but we argue here that the existence of large undeformed regions might instead arise naturally from a fractal style of continental faulting. A scaling analysis derived from laboratory simulations of the India–Asia collision^{2–5} yields a predicted distribution of undeformed regions, which can be tested in the field. We also show that, if continental faulting is fractal in nature, this will limit the validity of ‘homogenization’ approaches to the modelling of continental deformation, in which the rheology of the lithosphere is assumed to be homogeneous at scales larger than a threshold value.

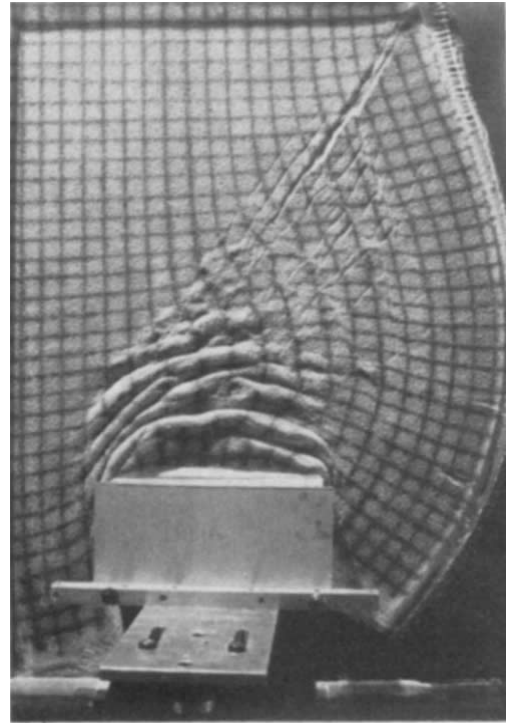
We have previously presented a series of experiments^{2–4}, scaled for gravity, on the formation of faults in a laboratory model of the Earth's lithosphere. The model respects the brittle–ductile nature of the lithosphere, and represents an attempt to model the India–Asia collision⁵, scaled down by a factor of 10^{-7} in length (1 m = 10,000 km) and $\sim 10^{-10}$ in time (1 h = 1 Myr). The presumed multilayer structure of the lithosphere is modelled as a sandwich^{5–7}: the upper brittle layer is made of uncemented quartz sand and floats on two lower layers, made of two silicone putties of different viscosity and density, simulating the ductile crust and a ductile mantle, respectively. The three layers float on a thick layer of golden syrup (corresponding to the asthenosphere) which ensures a free boundary condition at the bottom of the silicon layers. A rectangular wedge penetrates the system at a constant velocity.

As shown in Fig. 1, the simulation produces buckling (corresponding to the Himalayas) in front of the indenter, and penetrative deformation as far as 0.5 m (5,000 km) north of the indenter. Except in front of the indenter, the deformation generates a planar pattern of vertical faults deforming by horizontal shear.

Figures 2 and 3 show the results, in digitized form, of two experiments similar to that shown in Fig. 1, but with smaller ductile-layer viscosities, leading to weaker coupling between the ductile and brittle layers. In both cases, there is a very broad size distribution of fracture zones. Figure 2 shows two fault patterns corresponding to two different experiments performed under the same conditions (same layering, same rheologies and identical boundary conditions). The structure of the two fault patterns is similar but differs in detail. The position and length distributions are not identical demonstrating the sensitivity to initial conditions (such as very small differences in the compaction of the sand and in the boundary conditions), which is characteristic of a nonlinear process⁸ (here the stress-softening mechanism associated with faulting).

In a previous paper² we argued that the fault patterns are self-similar, with a fractal dimension $\Delta_f = 1.70 \pm 0.05$ in the interval 1–40 cm. Mathematically, this can be described by the number $N(l, r) \Delta l$ of faults of length between l and $l + \Delta l$ in a box of size r , which is found of the form $N(l, r) \Delta l = C r^b l^{-a} \Delta l$ where b is the fractal dimension of the fault barycentres, a is an exponent measured in Fig. 3c and C is a normalization constant. If $l_{\min}$ is the smallest fault size detected and $N_t(l_{\min})$ the total number of faults larger than $l_{\min}$ in the system of size Λ , $N(l, r)$ is normalized by the condition $\int_{l_{\min}}^{\Lambda} N(l, r = \Lambda) dl = N_t$, which

FIG. 1 Top view of a typical experiment and its set-up. The rectangular wedge, 20 cm wide, penetrates the system (40 × 68 cm) from the south. The deformed array of white passive markers was initially a perfect square lattice. The free boundary on the right was initially a straight line 10 cm away from the right side of the indenter. Here the rectangular parameters of this experiment were: upper brittle layer (sand), 0.5 cm thick, density = 1.2 g cm⁻³, upper layer of silicon 0.5 cm thick, density = 1.2 g cm⁻³, viscosity = 2.3 × 10⁴ Pa s; lower layer of silicon, 1 cm thick, density = 1.4 g cm⁻³, viscosity = 11 × 10⁴ Pa s; golden syrup, 5 cm thick, density = 1.47 g cm⁻³, viscosity = 100 Pa s.



yields the relation $N_l = \{C/(a-1)\} \Lambda^b l_{\min}^{-(a-1)}$. As $N(l, r) \Delta l$ is independent of $l_{\min}$ as long as $l > l_{\min}$, C should also be independent of $l_{\min}$. This implies that N_l scales as $l_{\min}^{-(a-1)}$. Given $l_{\min}$ (and thus N_l), $N(l, r)$ can be rewritten as

$$N(l, r) = (a-1) l_{\min}^{-1} N_l (r/\Lambda)^b (l_{\min}/l)^a \quad (1)$$

neglecting terms of order $(l_{\min}/\Lambda)^{a-1}$. $L(r)$ (Figs 2c, 3b), the total length of all faults within a disk of radius r , is obtained from equation (1) by

$$L(r) = \int_{l_{\min}}^r l N(l, r) dl + \int_r^\Lambda r N(l, r) dl \quad (2)$$

For $a \geq 2$, as seems to be the case in our experiments, the first term dominates, yielding $L(r) = \{C/(a-2)\} l_{\min}^{2-a} r^b$ for $r \gg l_{\min}$. Thus, the fractal dimension D_f , defined by the scaling law $L(r) \approx r^{D_f}$, should be equal to the barycentre exponent b and should be independent of the precise fault length distribution.

For $a < 2$, $D_f = b + 2 - a$ is a combination of a and b . We have checked these results by directly determining the fault barycentre exponent b , that is, by counting the number of fault centres in a box of size r as a function of r . We find that D_f is indeed close to b (ref. 9, and Ph.D., A.S. and D.S., manuscript in preparation), as it should be for $a \geq 1$. Equation (1) is the assumption on which the following discussion is based.

Our experiments allow us to address the question of the preservation of large undeformed regions in continental deformation. One spectacular example is the Tarim basin in the north of the Tibetan plateau, interpreted as an old craton that remained undeformed during the India-Asia collision. As can be observed in Figs 2a, b and 3a, the fractal structure of fault patterns is intimately connected with the existence of a hierarchy of undeformed domains of size ranging from $\sim l_{\min}$ to a fraction of the system size Λ . Could the existence of large undeformed areas be the result of the fractal fault pattern, with no need for

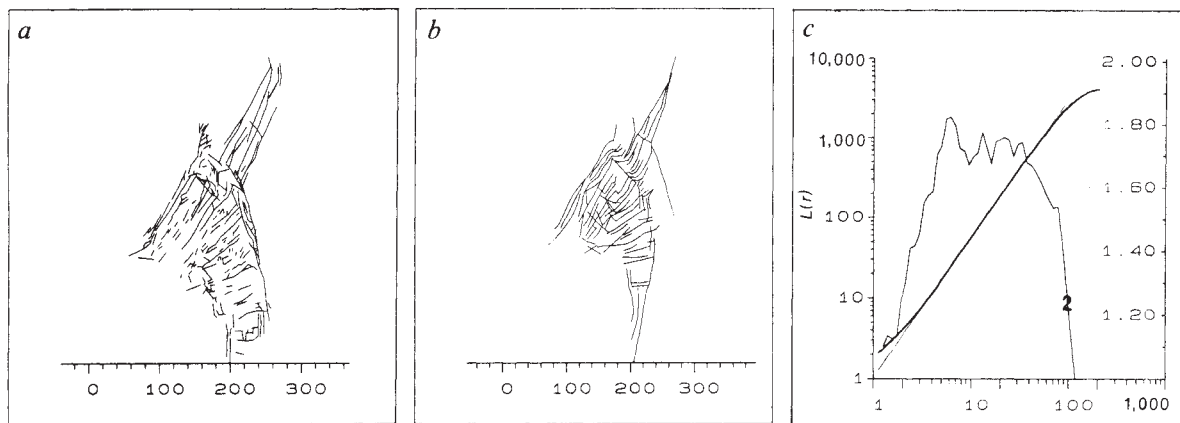


FIG. 2 a, b, Digitized fault patterns from two identical experiments 1a, b. Parameters as in Fig. 1 except that the viscosity of the lower layer of silicon is 4.5×10^4 Pa s. In both experiments the rectangular wedge has penetrated 25 cm into an initially undeformed system. c, $L(r)$ as a function of r . $L(r)$ is fitted by $L(r) \approx r^{D_f}$ (thin line) with $D_f = 1.73 \pm 0.05$. Both experiments give

the same value. The exponent a of the fault length distribution is 2.3 ± 0.1 . The curve labelled 2 gives the local value of the slope (right-hand scale) of the $\log L(r) - \log r$ plot. It is an estimate of the uncertainty of D_f . The correlation coefficient is > 0.99 for both experiments.

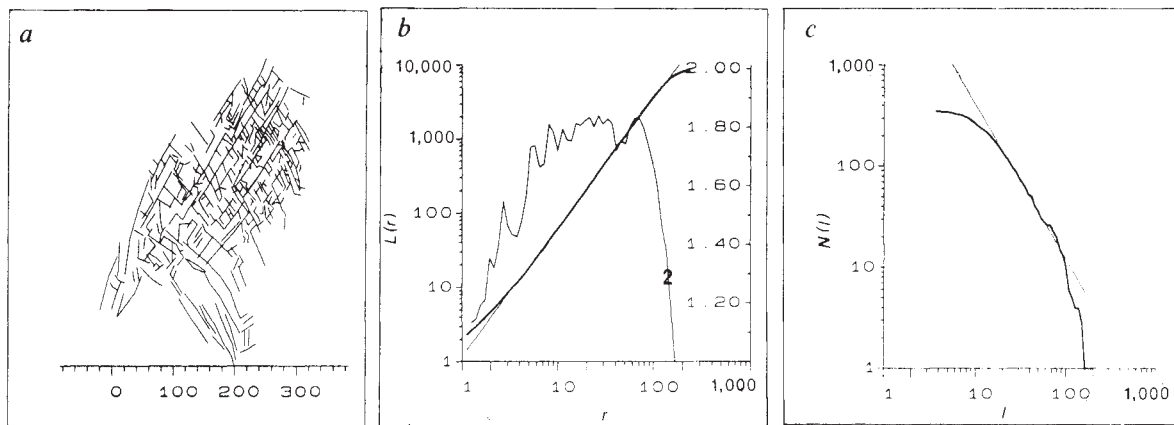


FIG. 3 a, Digitized fault pattern from experiment 2. The parameters were the same as in Fig. 1 except that the lower layer of silicon was 0.5 cm thick and its viscosity was 7×10^4 Pa s. The rectangular wedge has penetrated 15 cm into an initially undeformed system. b, $L(r)$ as a function of r for

faults in a box of size $r \times r$. As in Fig. 2b, curve 2 is the local slope. $L(r)$ is fitted by $L(r) \approx r^{D_f}$ (thin line) with $D_f = 1.73 \pm 0.05$. c, The number of faults $N(l) = \int_l^\infty P(l') dl'$ of length $> l$ is fitted by $N(l) \approx l^{-a+1}$ with $a = 2.5 \pm 0.1$.

macroscopic variations of strengths? To test this hypothesis, we used the above scaling law [equation (1)] to predict the distribution of stable domains areas A and their maximum size $A_{\max}(\Lambda, N_t)$.

For a domain D of size L and area $A = L^2$, the average number of faults in D of all lengths is, from equation (1), $N_t(L/\Lambda)^b$. Because the formation of fault pattern in our model is sensitive to low levels of noise stemming from various physical causes (experiments 1a and 1b differ although prepared under the same macroscopic conditions), we assume that the faults are randomly distributed according to the fractal distribution given by equation (1). This may not be entirely true, but this assumption is the least stringent in the absence of more information on the fault structure. For a random distribution, one expects fluctuations in the local density of faults. Within this markovian hypothesis, the probability $P(A)$ that the domain D contains no faults is

$$P(A = L^2) \approx [N_t(L/\Lambda)^b]^{-1/2} \exp\{-N_t(L/\Lambda)^b\} \quad (3)$$

The maximum stable domain size $A_{\max}(\Lambda, N_t)$ is obtained from the condition

$$P(A_{\max}) (\Lambda^2/A_{\max}) \approx 1 \quad (4)$$

which states that there is at least one domain of size $A_{\max}$ out of a total of $\Lambda^2/A_{\max}$ independent events, each of probability $P(A_{\max})$. Using equation (3) in equation (4), we obtain the leading behaviour

$$A_{\max}(\Lambda, N_t)/\Lambda^2 \approx \{[(b+2)/2b](\log N_t)/N_t\}^{2/b} \quad (5)$$

In our case, $b = D_f \approx 1.7$. Then, $A_{\max}(\Lambda, N_t) \approx L_{\max}^2$ is a slowly decreasing function of N_t : for $N_t = 10^2, 10^3, 10^4, 10^5, 10^6$, $L_{\max}/\Lambda \approx 0.22, 0.07, 0.02, 0.006, 0.002$. These results are not greatly sensitive to the precise value of b .

This calculation shows that the notion of an undeformed domain depends on the total number N_t of faults taken into account. It is clear that $N_t \approx l_{\min}^{-(a-1)}$ depends strongly on the size resolution $l_{\min}$ of the smallest faults taken into account. As faults can be observed from the scale of micrometres to thousands of kilometres, there is apparently no sense in numbering them. Our point is that a given region will qualify as undeformed if it does not contain any fault larger than some threshold $l_{\min}$ chosen for its convenience. A compilation of faults in the Asian continent has been performed⁵ for scales larger than $l_{\min} \approx 100$ km and $N_t(l_{\min})$ was found to be of the order of several hundreds. At this resolution, the Tarim basin seems to be the greatest domain that does not contain any fault greater than $l_{\min}$. Its size compares well with our estimation $L_{\max}/\Lambda \approx 0.1-0.2$ with $\Lambda \approx 3,000$ km. This suggests that, if continental faulting is indeed

fractal in nature, an explanation based on macroscopic strength variations is not required to explain the occurrence of large stable basins.

The scaling in equation (1) also allows us to analyse the consequence of fractal fault geometry on homogenization procedures. In these approaches, one introduces a representative volume with a length scale L_H . At scales larger than L_H , the system behaves as if it was homogeneous with effective average rheological laws. Owing to the importance of faults on the deformation field, this approach is certainly valid if no faults of size of the order of or larger than L_H cut the representative volume element in two pieces. Using equation (1), we can estimate the average number of faults $N(l \geq L_H)$ (which may be less than one) of size $l \geq L_H$ the barycentres of which lie on a surface of size L_H^2 . $N(l \geq L_H)$ quantifies the degree of heterogeneity of the representative element. From equation (1), we have

$$N(l \geq L_H) = N_t (l_{\min}/\Lambda)^{a-1} (L_H/\Lambda)^{b+1-a} \quad (6)$$

If $N(l \geq L_H)$ decreases as L_H increases, then some sufficiently large L_H qualifies as the homogenization length because $N(l \geq L_H)$ becomes less than one. This occurs for $b+1-a < 0$. If $b+1-a \geq 0$, $N(l \geq L_H)$ increases with L_H . This corresponds to an increasing heterogeneity at larger and larger observation lengths. In this case, homogenization cannot hold. Our experiments give $b \approx 1.7$ and $a \approx 2.1-2.5$ and therefore $b+1-a > 0$. We note that for very high viscosities of the ductile layers, the exponent a seems to increase above 2.5, making possible a transition to the regime $b+1-a < 0$. This correlates well with the denser morphology of the fault pattern associated with the stronger influence of the ductile smoothing rheology. Great care should be taken before using the homogenization approach to describe systems that could present fractal fault patterns. Further analysis of continental faults is needed to demonstrate the relevance of our analysis to nature. □

Received 28 March; accepted 3 September 1990.

1. Vilotte, J. P., Daignière, M., Madariaga, R. & Zienkiewicz, O. C. *Phys. Earth planet. Inter.* **36**, 236-259 (1984).
2. Sornette, A., Davy, P. & Sornette, D. *Phys. Rev. Lett.* (in the press).
3. Davy, Ph. & Cobbold, P. R. *Bull. geol. Instn. Univ. Uppsala* **14**, 129-142 (1988).
4. Davy, Ph. & Cobbold, P. R. *Tectonophysics* (in the press).
5. Cobbold, P. R. & Davy, Ph. *Bull. geol. Instn. Univ. Uppsala* **14**, 143-162 (1988).
6. Kirby, S. H. *Rev. Geophys. Space Phys.* **21**, 1458-1487 (1983).
7. Kirby, S. H. *Tectonophysics* **119**, 1-27 (1985).
8. Bergé, P., Pomeau, Y. & Vidal, C. in *Order within Chaos* (ed. Hermann) (Wiley, Paris, 1984).
9. Sornette, A. thesis, Univ. Orsay (1990).

ACKNOWLEDGEMENTS. We thank P. England and C. Vanneste for discussions. Contribution CNRS-INSU-DBT No. 185.

Extreme isotopic variations in Heard Island lavas and the nature of mantle reservoirs

Jane Barling*† & Steven L. Goldstein*

* Max-Planck-Institut für Chemie, Postfach 3060, D-6500 Mainz, Germany
† Department of Earth Sciences, Monash University, Clayton, Victoria 3168, Australia

OCEANIC basalts are produced by melting of the Earth's mantle and are widely used to probe its composition. The observation of systematic, coupled variations in the neodymium, strontium and lead isotopic compositions of these basalts¹ has been widely interpreted as reflecting the mixing of identifiable mantle components. White² and Zindler and Hart³ have suggested that the isotopic compositions of mid-ocean-ridge and ocean island basalts (MORB and OIB) can be considered as mixtures of a depleted MORB-type mantle component with a few other components reflecting long-term enrichment of Rb/Sr, Nd/Sm and/or U/Pb ratios. It remains unclear whether these components should be considered as hypothetical mixing endmembers, or whether they exist physically as mantle reservoirs. This question may be addressed if the mixing endmembers of individual mantle plumes can be identified. To this end, we report linear Pb–Pb and curvilinear Pb–Nd and Pb–Sr isotopic covariations in Recent lavas of Heard Island, southern Indian Ocean, indicating binary mixing. The hyperbolic relationships are unique among oceanic basalts and tightly constrain the isotopic compositions of the plume source components, which do not coincide with the mantle endmember compositions of ref. 3. If our results are generally applicable to other plumes, they call into question the existence of large mantle reservoirs corresponding to these components, and indicate that actual oceanic basalt source reservoirs have intermediate isotopic compositions.

Heard Island (53° 06' S, 73° 30' E), on the Kerguelen Plateau, is dominated by Big Ben Volcano (2,745 m) and peripheral cones on the Laurens Peninsula. The data reported here are part of a larger geochemical study (J.B., S.L.G. and I. A. Nicholls, manuscript in preparation) and display the largest range of Sr, Nd and Pb isotopic compositions yet observed for lavas erupted over a short time (<1 Myr) on an ocean island (for example, $^{87}\text{Sr}/^{86}\text{Sr} = 0.7047\text{--}0.7079$). Laurens Peninsula lavas range from basanite to trachyte and are isotopically homogeneous, whereas the predominantly basaltic and trachybasaltic Big Ben lavas are isotopically heterogeneous (Table 1). Isotopic compositions correlate with some trace-element ratios, for example, $^{87}\text{Sr}/^{86}\text{Sr}$ correlates negatively with Nb and Eu anomalies in samples in which plagioclase is not a crystallizing phase (J.B., S.L.G. and I. A. Nicholls, manuscript in preparation). The major-element and isotopic compositions do not correlate, however, indicating that the isotope and trace-element variations are not a result of shallow contamination of differentiating magmas, but rather reflect mantle-source characteristics. The lavas show the distinctive 'Dupal' isotope signatures^{4,5} typical of Indian Ocean basalts.

The curvilinear arrays of isotopic composition (Fig. 1) are compatible with two-component mixing in a mantle plume, between a 'depleted' Heard component, defined on the basis of having the lower $^{87}\text{Sr}/^{86}\text{Sr}$ and higher $^{143}\text{Nd}/^{144}\text{Nd}$ ratios (it also has the higher $^{206}\text{Pb}/^{204}\text{Pb}$ ratios), and a complementary 'enriched' Heard component. The curvature of the Pb–Sr and Pb–Nd arrays displayed by this suite is unique among oceanic basalts. They allow identification of the isotopic compositions of the endmembers, which we equate with the mantle-source components contributing to the lavas, and support a significant role for source mixing processes in generation of ocean island basalts. Isotopic compositions of Heard lavas are broadly similar to those of Kerguelen which do not display the curvilinear isotope arrays. In the following discussion, endmember refers

to the most extreme isotopic compositions of a data set, component refers to distinctive isotopic compositions in the mantle but without any physical implications, and reservoir refers to large physical masses having distinctive isotopic compositions.

The Pb–Sr and Pb–Nd isotope mixing curves (Fig. 1b, c) are derived from least-squares regressions of the data fitted to a hyperbola of the form $c = (y - y_0)(x - x_0)$, where y_0 and x_0 are the asymptotes, and c is a constant governing the curvature⁶. The Pb–Sr isotope array (Fig. 1) shows the greatest curvature and places the tightest constraints on the endmembers, yielding precise asymptotes of $(^{206}\text{Pb}/^{204}\text{Pb})_0 = 17.67 \pm 0.08$ and $(^{87}\text{Sr}/^{86}\text{Sr})_0 = 0.7044 \pm 0.0002$ (Fig. 1). Although the Pb–Nd data

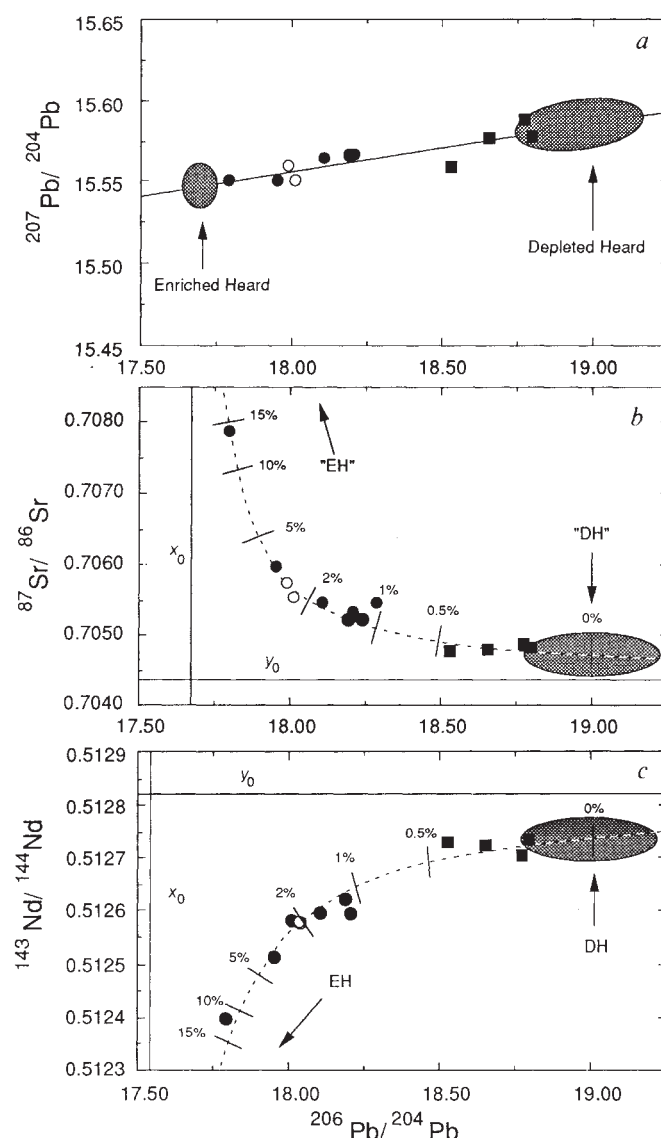


FIG. 1 a, $^{206}\text{Pb}/^{204}\text{Pb}$ plotted against $^{207}\text{Pb}/^{204}\text{Pb}$. The best-fit line is shown. b, c, $^{206}\text{Pb}/^{204}\text{Pb}$ plotted against $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{143}\text{Nd}/^{144}\text{Nd}$, respectively. Dashed lines are the best-fit hyperbolic curves, $(x - x_0)(y - y_0) = c$. For Pb–Sr, $(^{206}\text{Pb}/^{204}\text{Pb})_0 = 17.67$ (8), $(^{87}\text{Sr}/^{86}\text{Sr})_0 = 0.7044$ (2), $c = 0.000439$ (217). For Pb–Nd, $(^{206}\text{Pb}/^{204}\text{Pb})_0 = 17.54$ (32), $(^{143}\text{Nd}/^{144}\text{Nd})_0 = 0.51282$ (10), $c = -0.000122$ (140) (errors are 2σ). The field of the enriched Heard (EH) component is estimated from the Pb–Sr and Pb–Nd curves; the depleted Heard (DH) field from the covariation of $^{206}\text{Pb}/^{204}\text{Pb}$ – $^{87}\text{Sr}/^{86}\text{Sr}$ ratios in Indian Ocean basalts (Fig. 2). Tick marks show the amount of EH component in a mixture described by the best-fit curves and assuming $(^{87}\text{Sr}/^{86}\text{Sr})_{\text{EH}} = 0.710$, $(^{206}\text{Pb}/^{204}\text{Pb})_{\text{DH}} = 19.0$, $\text{Nd}_{\text{EH}} = 25$ p.p.m. and $\text{Sr}_{\text{DH}} = 20$ p.p.m. Squares and circles represent Laurens Peninsula and Big Ben samples, respectively. Filled symbols are from this study, open ones are from ref. 15. The $^{206}\text{Pb}/^{204}\text{Pb}$ – $^{208}\text{Pb}/^{204}\text{Pb}$ array (not shown) is also linear.

TABLE 1 Isotopic compositions of Heard Island lavas

Sample number	Location	SiO ₂ (%)	MgO (%)	Sr (p.p.m.)	⁸⁷ Sr/ ⁸⁶ Sr*	⁸⁷ Sr/ ⁸⁶ Sr†	¹⁴³ Nd/ ¹⁴⁴ Nd*	¹⁴³ Nd/ ¹⁴⁴ Nd†	²⁰⁶ Pb/ ²⁰⁴ Pb	²⁰⁷ Pb/ ²⁰⁴ Pb	²⁰⁸ Pb/ ²⁰⁴ Pb
65002 (TR)	LP	61.11	0.77	670	0.704778 (9) 0.704790 (15)	0.704784 (16)	0.512735 (10) 0.512727 (13)	0.512731 (9)	18.527	15.558	38.944
65054 (TA)	LP	50.56	4.09	1,189	0.704802 (16) 0.704805 (22)	0.704804 (14)	0.512728 (17) 0.512711 (17)	0.512720 (14)	18.656	15.577	38.980
65015 (TR)	LP	59.10	0.84	727	0.704817 (10) 0.704813 (26)	0.704815 (13)	0.512685 (14) 0.512712 (11) 0.512734 (6) 0.512777 (15)	0.512725 (28)	18.796	15.578	39.120
69244 (BN)	LP	45.26	7.17	939	0.704875 (14) 0.704857 (13) 0.704852 (8) 0.704872 (12) 0.704855 (15)	0.704862 (10)	0.512697 (8) 0.512712 (16)	0.512705 (16)	18.776	15.588	39.170
69244/2 H10 (BN)	BB	42.64	13.26	955	0.704861 (10) 0.705233 (12) 0.705235 (20)	0.704861 (60) 0.705234 (12)	0.512625 (20) 0.512610 (16)	0.512618 (18)	18.189	15.566	38.646
65171 (TB)	BB	48.18	8.34	818	0.705319 (16) 0.705345 (12) 0.705346 (12)	0.705337 (15)	0.512593 (16) 0.512599 (18)	0.512596 (13)	18.211	15.567	38.508
65085 (B)	BB	45.49	15.93	589	0.705467 (13) 0.705468 (10)	0.705468 (8)	0.512569 (11) 0.512599 (15) 0.512601 (21) 0.512600 (9) 0.512603 (12)	0.512594 (14)	18.110	15.564	38.590
65151 (TB)	BB	49.56	8.48	604	0.705993 (13) 0.706001 (10) 0.705994 (20)	0.705996 (9)	0.512519 (12) 0.512509 (17)	0.512514 (12)	17.953	15.550	38.420
69285 (TR)	BB	60.64	1.11	611	0.707897 (14) 0.707976 (14) 0.707898 (12) 0.707905 (8)	0.707921 (39)	0.512377 (30) 0.512396 (14)	0.512387 (17)	17.790	15.551	38.358
69285/2					0.707979 (16) 0.707965 (26)	0.707972 (16)					

Nd, Sr and Pb isotopes were measured at MPI-Mainz. Chemical techniques are slightly modified from refs 26, 27. Measurements were by static (most Nd, Sr, Pb) or dynamic (some Nd) multicollection on two MAT 261 mass spectrometers. Sr and Nd: All samples were leached for ≥ 1 h in hot 6N HCl. ⁸⁷Sr/⁸⁶Sr are normalized to ⁸⁶Sr/⁸⁸Sr = 0.1194 and corrected to ⁸⁷Sr/⁸⁶Sr = 0.71023 based on measurements on NBS987 (⁸⁷Sr/⁸⁶Sr = 0.710212 (5) (where the number in parentheses is the 2σ error on the mean, $n=12$, $\sigma=9$) and 0.710240 (15) ($n=16$, $\sigma=30$) during different periods). ¹⁴³Nd/¹⁴⁴Nd are normalized to ¹⁴⁶Nd/¹⁴⁴Nd = 0.7219 and corrected to ¹⁴³Nd/¹⁴⁴Nd = 0.511860 based on measurements of the La Jolla standard (¹⁴³Nd/¹⁴⁴Nd = 0.511846 (6) ($2\sigma_m$, $n=15$, $\sigma=12$) (static) or 0.511842 (13) ($n=9$, $\sigma=19$) (dynamic)). Most samples were measured at least twice. Individual mass spectrometer runs and in-run precision are given. Averages are also given, with errors are expressed as $2\sigma_m$ —these errors are greater than the deviation of the means of the individual runs because they also take into account the in-run errors. In most cases the reproducibility of samples is better than that of the standards. Samples 69244/2 and 69285/2 are separate dissolutions. Pb: All (powdered) samples were leached overnight in cold 1N HNO₃. Pb isotopes are corrected relative to NBS values of standard NBS982 by factors of 0.154% a.m.u.⁻¹ for ²⁰⁷Pb/²⁰⁴Pb and ²⁰⁸Pb/²⁰⁴Pb and 0.179% a.m.u.⁻¹ for ²⁰⁶Pb/²⁰⁴Pb, accounting for fractionation and a reproducible cup bias on ²⁰⁶Pb. Reproducibility of NBS982 was better than 0.06% a.m.u.⁻¹ (2σ , $n=12$). Pb isotopes were measured at least twice and results are averages. Total preparation blanks were <400 pg. SiO₂ and MgO were measured by X-ray fluorescence at Monash University, except for samples 69244 and 69285 which were analysed at the University of Tasmania. Sr was measured by isotope dilution at MPI-Mainz, except for samples 65002, 65054 and H10 which were analysed by X-ray fluorescence at Monash University.

* Results of individual mass spectrometer runs from the same dissolution.

† Combined results based on statistical average of individual runs. BB, Big Ben; LP, Laurens Peninsula; BN, basanite; B, basalt; TB, trachybasalt; TA, trachyandesite; TR, trachyte.

are more scattered, the asymptote values of (¹⁴³Nd/¹⁴⁴Nd)₀ = 0.51282 ± 0.00010 and (²⁰⁶Pb/²⁰⁴Pb)₀ = 17.54 ± 0.32 are nevertheless well constrained. The asymptotes limit the maximum Nd and minimum Sr isotopic ratios of the depleted endmember, and the minimum ²⁰⁶Pb/²⁰⁴Pb ratio of the enriched endmember. The Sr and Nd asymptote values relevant to the depleted endmember are much more enriched than Indian Ocean MORB values, and also more enriched than PREMA, which was suggested by Zindler and Hart³ to be the depleted component of most OIB (Fig. 2), and which has isotopic compositions more enriched than MORB.

The identity of the Heard Island plume's enriched endmember can be evaluated from the mixing curves. Vollmer⁷ and Langmuir *et al.*⁸, using the hyperbola equation in the form $Ax + Bxy + Cy = -D$, showed that a factor r , related to the element ratios of the endmembers, can be calculated. For example, $r_{\text{PbSr}} = [(Pb/Sr)_{\text{EH}}/(Pb/Sr)_{\text{DH}}]$, where EH and DH are the enriched and depleted Heard endmembers, respectively. It can be shown that the value of r varies with the values of the mixing endmembers according to the relationship $r = (y_2 - y_0)/(y_1 - y_0) = (x_1 - x_0)/(x_2 - x_0)$, where (x_1 , y_1) and (x_2 , y_2) in this example refer to the depleted and enriched endmembers, respectively. This equation shows that as (⁸⁷Sr/⁸⁶Sr)_{EH} or (²⁰⁶Pb/²⁰⁴Pb)_{DH} increases (Fig. 1b), r_{PbSr} also increases. Because the true mixing endmembers of the Heard plume are likely to have isotopic compositions beyond the data set, using the Heard lavas having the most extreme Sr and Pb isotopic compositions (Table 1) as endmembers gives minimum estimates of r_{PbSr} (= 10) and r_{PbNd} (= 5). Thus the enriched Heard endmember must have

much higher Pb/Sr, Pb/Nd and Nd/Sr ratios than the depleted endmember, relationships typical of continental crust relative to mantle compositions⁹⁻¹¹.

These observations are very robust, depending only on the assumption of binary mixing (Fig. 1). Other possible mixing models, for example, that a mixing component has isotopic compositions of depleted Big Ben lavas, will be discussed elsewhere (J.B., S.L.G. and I. A. Nicholls, manuscript in preparation). In all mixing models the first-order implications are the same. First, direct contributions from Indian Ocean MORB or PREMA sources must be so small as to have insignificant effect on isotopic compositions of Heard lavas. The lack of involvement of these sources cannot be deduced from Sr–Nd isotope array because of its small curvature, which points toward MORB at the low ⁸⁷Sr/⁸⁶Sr end (Fig. 2a). Second, the enriched Heard component has high Pb/Sr, Pb/Nd and Nd/Sr, as well as high ⁸⁷Sr/⁸⁶Sr and low ¹⁴³Nd/¹⁴⁴Nd ratios compared with the depleted component. These isotope and trace-element features, together with the anticorrelation of ⁸⁷Sr/⁸⁶Sr with Nb and Eu anomalies (J.B., S.L.G., and I. A. Nicholls, manuscript in preparation), and combined with evidence that the enriched component is a mantle-source feature, indicate that the enriched Heard endmember is derived from upper continental crust that has been recycled to the mantle, supporting previous suggestions concerning the nature of Dupal components in Indian Ocean OIB^{2,4,12-14}. The continental lithospheric mantle has also been suggested as the source of enriched components in the Indian Ocean mantle^{15,16}. Analyses of lithospheric mantle xenoliths have shown neither low Sr/Nd ratios nor depletions in Nb and

Eu (ref. 17). A more definitive characterization of the composition of the continental lithospheric mantle is needed in order to distinguish finally between the possibilities; in the interim, our guess is that the enriched component is peridotitic mantle that has been modified by sediment subduction.

Our conclusions do not depend on the outlying sample (69285, Table 1), which is a trachyte. The same isotope and trace-element covariations, including curvature of the Pb–Sr and Pb–Nd isotope arrays, are displayed by the rest of the suite. If this sample is excluded, then redetermining the best-fit hyperbolae and assuming $(^{87}\text{Sr}/^{86}\text{Sr})_{\text{EH}} = 0.7060$ gives $r_{\text{PbSr}} = 8$. As a result of error magnification owing to the close approach of $(^{143}\text{Nd}/^{144}\text{Nd})_0$ to the data, r_{PbNd} is very sensitive to the value used for $(^{143}\text{Nd}/^{144}\text{Nd})_{\text{EH}}$, but it is still much greater than one. In any case, the isotopic compositions of the outlying sample lie on the Heard isotope arrays (Fig. 1) and seem to reflect mixing of the same two components as the rest of the suite.

The isotopic compositions of the Heard mixing endmembers have important implications for the nature of mantle magma reservoirs. Although small enriched 'blobs' or 'plums' are likely to exist in MORB sources and are sometimes sampled¹⁸, the absence of a MORB component in the Heard magma sources indicates that they are isolated from the shallow MORB mantle, suggesting that Heard and MORB sources exist as distinct mantle reservoirs. The segregation of OIB and MORB sources seems to be a general feature of at least the Indian Ocean mantle and is indicated by the independent regional Pb–Sr isotope arrays of Indian Ocean MORB and OIB (Fig. 2). The presence of components of continental origin in some OIB (this study and refs 12, 19–21), coupled with the apparent absence of MORB contributions to Heard lavas, shows that continental material is transported to the deep OIB source regions. Subduction provides a suitable mechanism for transporting continental crust into the mantle, and our study lends support to the suggestion that OIB are generated from subducted oceanic crust²², although other models allowing recycling of continental crust to the mantle^{23,24} cannot be excluded. The contribution of continental crust to the Heard Island source is small (<3.5 wt%, assuming $(^{87}\text{Sr}/^{86}\text{Sr})_{\text{EH}} \geq 0.7100$) for all samples except the lava having the most extreme isotopic compositions (Fig. 1b, c). Small contributions also apply to other OIB suites²⁵.

Zindler and Hart³, showed that all oceanic basalts can be described as mixtures of four components, (subsequently referred to as ZH components), called DMM (depleted MORB mantle), HIMU (having high $^{206}\text{Pb}/^{204}\text{Pb}$ and low $^{87}\text{Sr}/^{86}\text{Sr}$ ratios), EM1 and EM2 (enriched mantle components having low $^{143}\text{Nd}/^{144}\text{Nd}$ ratios, but with EM1 having lower $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{206}\text{Pb}/^{204}\text{Pb}$ ratios) (Fig. 2b). Because these components are considered as endmembers in the oceanic mantle, their isotopic compositions are defined to be outside of the range of oceanic basalts (although it has since been suggested that isotopic compositions of HIMU coincide with Mangaia Island¹³). If the ZH components exist in the mantle as reservoirs that mix before eruption of basalts, then masses of material having isotopic compositions of ZH components must be widespread. If, on the other hand, ZH components are hypothetical mixing endmembers, there is no need for the large-scale existence of mantle reservoirs having their isotopic compositions. The Heard Island data are significant in this regard because they closely constrain the isotopic compositions of mixing endmembers of a plume and provide the best indications currently available of the nature of the source reservoirs of oceanic basalts. They demonstrate the importance of mixing in the generation of oceanic basalts but the mixing components in Heard's sources have isotopic compositions of neither the ZH endmember components nor PREMA.

The isotopic composition of enriched Heard component lies between that of EM1 and EM2 (Fig. 2), indicating either that (1) it is a mixture of EM1 and EM2, or (2) it is a distinct EM component (EM3), or (3) EM components do not have distinct

isotopic compositions, but rather EM represents a spectrum of isotopic compositions between the most extreme compositions represented by EM2 and EM1. Mantle contaminated by continental crust should display a large range of isotopic compositions. To the extent that EM2 is ultimately derived from continental crust, the enriched Heard component should be considered as a subset of EM2, extending the range of EM2 isotopic compositions, and blurring the distinction between EM2 and EM1. We favour case (3) above and suggest that EM compositions reflect mantle contaminated by recycled continental crust.

The depleted Heard component has isotopic compositions intermediate to all of the ZH components. Assuming intermediate compositions are typical of ocean island sources, this indicates that (1) ZH components are mixed to intermediate compositions before formation of plumes, or (2) ZH components have relatively minor roles as reservoirs. Zindler and Hart³ suggest that oceanic basalt sources may reflect intimate mixtures of ZH components on size scales smaller than that sampled by melting. Case (2) above, however, does not require the widespread existence of material having isotopic compositions of ZH components on any size scale.

The success of ZH components in describing oceanic basalts indicates that MORB and OIB fall into groups having similar time-integrated histories. EM compositions in OIB seem to be derived from material with long (>10⁹ yr) histories outside the convecting mantle (in the continental crust, for example). Other compositions probably reflect the competing processes of

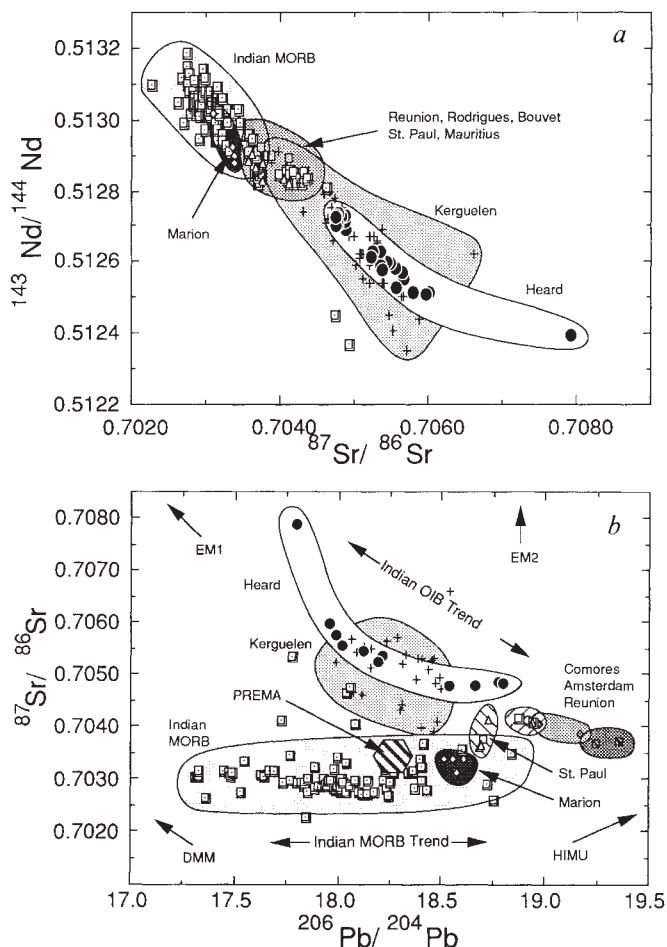


FIG. 2 Isotopic compositions of Indian Ocean MORB and OIB. a, $^{87}\text{Sr}/^{86}\text{Sr}$ plotted against $^{143}\text{Nd}/^{144}\text{Nd}$. b, $^{206}\text{Pb}/^{204}\text{Pb}$ plotted against $^{87}\text{Sr}/^{86}\text{Sr}$. In both diagrams, black circles show the Heard data (from this study, J.B. S.L.G. and I. A. Nicholls, manuscript in preparation; ref. 15). Other data are from the literature, recently summarized in refs 13, 28. In Fig. 2b, DMM, HIMU, EM1, EM2 and PREMA are from ref. 3.

intramantle differentiation (including generation of oceanic crust, which has only a short history, $\sim 10^8$ yr, outside the mantle) and convective mixing, which together would tend to produce mantle reservoirs having isotopic compositions intermediate to ZH components. The lack of basalts having compositions near the DMM and EM components, and the small amount of basalt having HIMU compositions, do not lend support to the existence of these components as large mantle reservoirs. Rather, the few basalts having compositions close to ZH components may reflect the most fractionated examples or the longest-surviving remnants of mantle fractionation processes. □

Received 24 April; accepted 11 September 1990.

1. Zindler, A., Jagoutz, E. & Goldstein, S. *Nature* **298**, 519–523 (1982).
2. White, W. M. *Geology* **13**, 115–118 (1985).
3. Zindler, A. & Hart, S. R. *Rev. Earth planet. Sci.* **14**, 493–571 (1986).
4. Dupré, B. & Allègre, C. J. *Nature* **303**, 142–146 (1983).
5. Hart, S. R. *Nature* **309**, 753–757 (1984).
6. Juteau, M., Michard, A. & Albarède, F. *Contr. Miner. Petrol.* **92**, 331–340 (1986).
7. Vollmer, R. *Geochim. cosmochim. Acta* **40**, 283–295 (1976).
8. Langmuir, C. H., Vocke, R. D. Jr, Hanson, G. N. & Hart, S. R. *Earth planet. Sci. Lett.* **37**, 380–392 (1978).
9. Taylor, S. R. & McLennan, S. M. *The Continental Crust* (Blackwell Scientific, Oxford 1985).
10. Newsom, H. E., White, W. M., Jochum, K. P. & Hofmann, A. W. *Earth planet. Sci. Lett.* **80**, 299–313 (1986).
11. Hofmann, A. W., Jochum, K. P., Seufert, M. & White, W. M. *Earth planet. Sci. Lett.* **79**, 33–45 (1986).
12. Weaver, B. L., Wood, D. A., Tarney, J. & Joron, J.-L. *Geology* **14**, 275–278 (1986).
13. Hart, S. R. *Earth planet. Sci. Lett.* **90**, 273–296 (1988).
14. Gautier, I., Giret, A., Loubet, M., Vidal, P. & Weis, D. *Earth planet. Sci. Lett.* (submitted).
15. Storey, M. *et al.* *Nature* **336**, 371–374 (1988).
16. Storey, M. *et al.* *Nature* **338**, 574–576 (1989).
17. Jochum, K. P., McDonough, W. F., Palme, H. & Spettel, B. *Nature* **340**, 548–550 (1989).
18. Zindler, A., Staudigel, H. & Batiza, R. *Earth planet. Sci. Lett.* **70**, 175–195 (1984).
19. Palacz, Z. & Saunders, A. D. *Earth planet. Sci. Lett.* **79**, 270–280 (1986).
20. Dupuy, C., Vidal, P., Barszczus, H. G. & Chauvel, C. *Earth planet. Sci. Lett.* **82**, 145–152 (1987).
21. Devey, C. W. *et al.* *J. geophys. Res.* **95**, 5049–5066 (1990).
22. Hofmann, A. W. & White, W. M. *Earth planet. Sci. Lett.* **57**, 421–436 (1982).
23. McKenzie, D. & O'Nions, R. K. *Nature* **301**, 229–231 (1983).
24. Arndt, N. T. & Goldstein, S. L. *Tectonophysics* **161**, 201–212 (1989).
25. Sun, S.-S. & McDonough, W. F. in *Magma-tism in Ocean Basins* (eds Saunders, A. D. & Norry, M. J.) 313–345 (Geol. Soc. Lond., London, 1989).
26. White, W. M. & Patchett, P. J. *Earth planet. Sci. Lett.* **67**, 167–185 (1984).
27. White, W. M. & Dupré, B. *J. geophys. Res.* **91**, 5927–5941 (1986).
28. Mahoney, J. J. *et al.* *J. geophys. Res.* **94**, 4033–4052 (1989).

ACKNOWLEDGEMENTS. We thank F. Albarède, C. Chauvel, C. Devey, S. Galer, A. Hofmann and I. Nicholls for comments, S. Hart and A. Saunders for reviews, F. Albarède for sharing software, and G. Wheller for samples 69244 and 69285. J.B. acknowledges the Australian Antarctic Division and the Australian Antarctic Science Advisory Committee for field and financial support, and the Deutsche Akademische Austauschdienst for a scholarship.

Male mate choice as predicted by sperm competition in thirteen-lined ground squirrels

P. L. Schwagmeyer* & G. A. Parker†

*Departments of Psychology and Zoology, University of Oklahoma, Norman, Oklahoma 73019, USA

†Department of Environmental and Evolutionary Biology, University of Liverpool, Liverpool L69 3BX, UK

MALE mate choice is predicted to occur when males transfer nutrients to females during copulation, provide paternal care, or are otherwise limited in mating capacity^{1–7}. Most commonly, field studies have revealed male choice to be focused on females offering high fecundity^{8–10}. Here we describe male mating selectivity that appears to be driven by sperm competition: male thirteen-lined ground squirrels (*Spermophilus tridecemlineatus*) avoid copulating with previously mated females once the fertilization gain rate drops below the average obtainable from alternative matings. Although this behaviour seems advantageous to males, it limits the duration of female sexual activity.

In many polygynous species, male reproductive success hinges primarily on fertilizations achieved per unit time invested¹¹. For male thirteen-lined ground squirrels, pay-offs from mating can vary markedly owing to sperm competition. Oestrous females commonly mate with any males that find them and attempt to

TABLE 1 Model for discriminative mating with empirically derived variable estimates*

Fertilization gain rate for accepting an encountered female	Fertilization gain rate for rejecting an encountered female and resuming search
$E_2(d)/H_2(d)$	$E_{ov}/(S+P)$
$E_2(d)$: $-0.158d + 0.620$	E_{ov} : 0.433
$H_2(d)$: $-0.198d + 1.742$	S : 20.7
	P : 2.1

* See Table 2 for information on variation in the estimates.

$H_2(d)$ was calculated from data collected on 36 oestrous females in an Oklahoma population, where H was measured as the time elapsing between a second, third or fourth male's first mount and his departure, and d was the interval between termination of the first male's final copulation with a female and the time at which the 'delayed' male initially mounted. Handling time declined significantly with increased delay ($r^2 = 0.148$; $F = 5.92$, $P = 0.02$). $E_2(d)$: Estimates of the proportion of offspring sired were calculated for that same subset of field matings, involving males that were second, third or fourth to mate with a female. The estimates were based on laboratory results relating paternity to delay and copulation duration in a two-mates/female situation¹⁶. Linear regression was used to approximate the average decline in estimated paternity under field conditions as a function of delay. The regression excludes matings ($n = 9$) for which information on copulation duration or delay was incomplete, or that violated the laboratory result assumption of sequential male matings. Matings of third and fourth males were assumed to result in zero fertilization, an exclusion we base on: (1) four field matings in which electrophoretic paternity results indicated that there was little probability that they were effective in fertilization, and (2) one three-male laboratory mating that definitely yielded no offspring sired by the third male (see Table 2, second footnote). E_{ov} : Across four mating seasons¹³, the number of different mates per female has averaged 2.31 (s.e. = 0.138; $n = 42$ females). The mean proportion of a litter sired per male per mating (E_{ov}) was approximated as the reciprocal. P : The average time males devote to pursuit of and copulation with a female equalled 2.10 h (s.d. = 1.32; $n = 63$ matings). S : The mean search time necessary for location of an oestrous female was estimated as 20.7 h, based on observations of males during hours when at least one female in the population was in oestrus¹⁷.

copulate¹². Typically, each female mates with two males¹³; the first sires roughly 75% of the litter¹⁴. Ovulation in this species is induced¹⁵, and two factors are known to influence sperm competition outcomes: delay between male matings (longer delays favour first males), and duration of the second male's longest copulation (increases favour second males)¹⁶.

Male mating opportunities also differ in temporal costs. Male mating order usually corresponds with the order in which males locate an oestrous female: the first male to arrive has priority, and copulation by a later male is deferred until the first departs (the 'queueing' convention¹⁷). On average, being late to arrive costs a male 1.9 h in waiting time if a competitor is present. Nevertheless, an evolutionarily stable strategy model addressing whether late-arriving males should leave immediately to seek unoccupied females revealed that male search costs (that is, the time to find another oestrous female) are high enough to make waiting the better option¹⁷.

High search costs generally are expected to constrain male mating selectivity⁶. Yet thirteen-lined ground squirrel males sometimes eschew mating opportunities, even in the absence of competitors (that is, without increased time costs). Our study investigates whether this observed selectivity corresponds with that expected from the effects of sperm competition on fertilization rate. We present a simple quantitative model that estimates a threshold for male rejection of an oestrous female, which we then compare with field data on discriminative male mating.

We assume that if a male finds a female acceptable, he pursues and copulates, thereby incurring a handling time cost. From optimal foraging theory¹⁸, a male should pursue a female only if his gain per unit handling time exceeds the average gain rate from the whole system. Hence, for the male to pursue, instead

of leaving to search for another female, requires that

$$E_2(d)/H(d) \geq E_{ov}/(S+P)$$

Here $E_2(d)$ represents the estimated proportion of a litter sired by a male mating d hours ($d > 0$) after a first male has finished copulating, and $H(d)$ is his mating time investment. E_{ov} is the mean proportion of a litter sired per mating (including those of first males at zero delay), S represents the average search hours necessary to locate an oestrous female, and P equals the average time a male spends pursuing a female once he has found her. P consequently includes both the actual handling time of males that mate, plus any waiting time imposed by a competitor's presence (Table 1).

To calculate the critical delay beyond which rejection of mating is favoured, we solved for d after setting the expected gain rate from pursuit (Table 1): $(-0.158d + 0.620)/(-0.198d + 1.742)$ equal to the average gain rate: $0.43/(20.7 \text{ h} + 2.1 \text{ h})$. The predicted mean threshold¹⁹ for male rejection of mating opportunities (d_{crit}) occurs about 3.8 h after females finish copulating with their first mates (Table 2).

To assess how well male rejection behaviour conforms to the model's prediction, we classified male responses to previously mated oestrous females into clear cases of acceptance or apparent cases of rejection, and examined their distribution as a function of delay (Fig. 1). Of 34 females, 11 (32%) experienced at least one interaction that qualified as a rejection; ten of the 11 did not mate with any other male after being rejected. Rejections tended to increase with delay (Fig. 1), and logistic

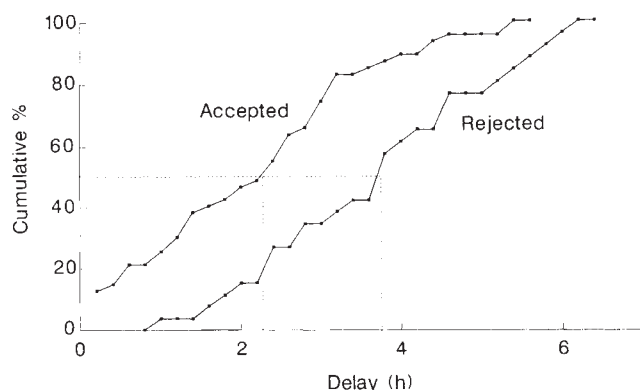


FIG. 1 Cumulative distributions of acceptances and rejections of previously-mated females as a function of delay (number h elapsed since the female finished copulating with her first mate). We considered a male clearly to have accepted a female if he pursued and mated with her ($n=47$ cases, 27 different males). Rejection was considered to have occurred if the male contacted an unaccompanied female (approaching within 2–3 m, or being approached by her), but did not attempt to mount her ($n=26$ cases, 13 different males). Note that rejections specifically exclude cases ($n=13$) that were complicated by the presence of a rival male. Some ($n=14$) were by males that already had mated with that female earlier in the day, whereas others ($n=12$) involved males that had not previously mated with the female. The value of mating should be roughly equivalent in both cases, except where a former mate encounters a female that has not re-mated since his departure. The latter circumstance occurred in only 8% of the mating opportunities. As shown, the median delay for rejection occurred at 3.81 h. Rejections occurred at significantly longer delays than acceptances (Kolmogorov-Smirnov 2-sample test, $X^2=11.09$, 2 d.f. $P<0.01$, one-tailed). To explore whether they might be accounted for more accurately by alternative factors that are also correlated with fertilization prospects, we used logistic regression analyses to compare the importance of delay as an independent variable with: (1) the number of different males each female had mated with (range 1–3); and (2) time of day (see Table 2, third footnote). Delay proved most useful in predicting whether males would reject (improvement $X^2=16.41$, 1 d.f., $P<0.001$). Nevertheless, because both number of mates and time of day are highly correlated with delay ($r=0.567$, $P<0.0001$ and $r=0.735$, $P<0.0001$, respectively), further research is necessary to determine the nature of the actual cue used by males as a mating criterion.

regression showed that rejection and acceptance became equally likely (predicted probability = 0.50 ± 0.08 s.e.) at 3.82 h delay, closely matching the 3.8 d_{crit} threshold derived from the model. Altogether, the model correctly predicted male behaviour during 74% (54/73) of the mating opportunities. The data further suggest that rejection of females at delays greater than 3.8 h did not impede the overall copulatory performance of males: the majority of >3.8 h 'rejectors' (8/9 that could be scored) achieved above-average mating success for the season.

It is possible, of course, that male rejections are due merely to waning of female attraction to males, rather than the reverse. Several lines of evidence suggest that this is not the case. First, we examined copulation termination data for females ($n=13$)

TABLE 2 Effects of variation in parameter estimates and seasonal fluctuations in variable values on d_{crit} outcome

Model component		Value	d_{crit}
$E_2(d)$ slope*	L1	-0.118	5.15
	L2	-0.198	3.03
$E_2(d)$ intercept	L1	0.729	4.52
	L2	0.512	3.11
$E_2(d)$ alternative†			
regression equation		$-0.176d + 0.727$	4.03
$H_2(d)$ slope	L1	-0.060	3.86
	L2	-0.336	3.74
$H_2(d)$ intercept	L1	1.367	3.86
	L2	2.117	3.76
$E_{ov} \ddagger$ Seasonal	min	0.34	3.83
	max	0.50	3.79
Diurnal	a.m.	0.59	3.76
	p.m.	0.26	3.85
$S + P \S$	min	19	3.78
	max	40	3.86
$E_{ov}/(S + P) \parallel$	min	0.012	3.84
	max	0.029	3.74

L1 and L2 represent the upper and lower limits of 90% confidence intervals. Because male search costs in this species are so high, the overall gain rate ($E_{ov}/(S+P)$) is very low, approaching zero. As a consequence, d_{crit} occurs only slightly before the value of mating drops to 0 ($E_2(d)=0$ when $d=3.92$), and it remains relatively stable in response to observed variation in E_{ov} , S and P . Only changes in male fertilization success as a function of delay ($E_2(d)$) have substantial effects.

* Confidence limits for the $E_2(d)$ parameter estimates are based on the error terms for the estimated paternity of field matings regression model. Note that earlier rejection would be favoured if paternity declined more rapidly with delay, or if delayed males had a greater initial fertilization disadvantage.

† A more conservative version of $E_2(d)$ was calculated ($E_2 = -0.176d + 0.727$; $n=16$), in which matings by third and fourth males were included in the sample and assigned zero paternity only if their combinations of delay and copulation duration would have yielded no paternity to a second male.

‡ The average number of mates per female has varied seasonally from a minimum of 2.0 in 1982 ($n=11$ females) to a maximum of 2.9 in spring of 1984 ($n=12$ females). Hence, the average proportion of a litter sired per male per mating (E_{ov}) has fluctuated between 0.345 and 0.500. Because nearly all females in this population begin their sexual activity in the morning, delay is highly correlated with time of day (Fig. 1). As a consequence, morning matings yield a higher estimated pay-off than afternoon matings. If male searches were random (that is, if the probability of encountering an oestrous female per unit search-time remains constant), males might be expected to shift their mating criterion during the day²⁴. But the daily decline in E_{ov} has little effect on d_{crit} .

§ From observations of male behaviour, we estimate that each male spends about 8.5 h per day of the mating season in mating-related activities (searching, waiting or mating). The average number of hours invested per male per mating ($S+P$) can therefore be estimated as $8.5 M \times L/(F \times C)$, where M is the number of resident males, L the number of days in which at least one female was sexually receptive, F the number of sexually active females, and C the average number of mates per female. The minimum and maximum estimates for four seasons are presented.

|| Calculated on a per-season basis, the minimum overall gain rate has been 0.012 litters per hour; the maximum, 0.029.

TABLE 3 Comparisons of events occurring during rejections ($n=26$) and acceptances

	Acceptance*	Rejection	P
Male or female aggression	1/17	3/26	>0.10†
Female ran from/ evaded male	11/17	17/26	>0.10‡
Female did not evade male:			
Male attempts mount	6/6	0/9	
Male leaves	0/6	8/9	
Total male approaches	1.88 (s.d.=1.27)	0.80 (s.d.=0.91)	<0.01§
Male approach rate (per min)	0.50 (s.d.=0.41)	0.21 (s.d.=0.29)	<0.01

* The acceptance cases considered here were those in which no competitor was present ($n=17$); male-female interactions before the male's first mount, which occurred an average of 8.9 (s.d.=12.91) min after his arrival, are compared with the interactions occurring before rejecting males departed from the female.

† $\chi^2=0.008$, 1 d.f.

‡ $\chi^2=0.079$, 1 d.f.

§ Mann-Whitney $U=100$, $z=3.00$.

|| Mann-Whitney $U=114$, $z=2.66$.

that mated with at least three males, to determine whether females became less receptive²⁰ with either increases in delay and/or sexual activity. Although third males began mating at an average delay of 3.2 h (s.d.=1.09), females terminated roughly the same percentage of copulations with them as they had with first males ($\bar{x}=81.1\%$ for first male copulations; 80.6% for third mates; analyses include only copulations unambiguously terminated by either participant; Mann-Whitney $U=53.5$; $P>0.10$). Second, we compared female behaviour during rejections and acceptances (Table 3), and found no differences in aggressiveness or evasiveness. Instead, rejections were clearly distinguishable by a lack of male persistence (Table 3). In contrast to the mean of 2.1 h that males spent pursuing females with which they mated, male visits with rejected females lasted only 6.0 min (s.e.=0.98). Rejected females often (15/26 cases) either reinitiated contact with a male by approaching him, or attempted to prolong contact by following him as he departed. Overall, our impression is that rejected females were still willing to mate, but were no longer sexually attractive²⁰.

According to recent arguments, the advantage to males of mating selectivity depends on two key features: variation in female quality and costs of obtaining alternative mates^{2,4,6}. As applied here to thirteen-lined ground squirrels, male choice should be favoured, even in the absence of paternal investment and despite high search costs, because of predictable variation in female quality arising from sperm competition. It seems possible that male choosiness, as mediated by cues regulating male sexual interest, may commonly be driven by sperm competition and, in turn, may frequently limit the duration of behavioural oestrus in female mammals. A lack of male attraction to females at particular stages of oestrus (as revealed by either male-initiated termination of pursuit or male refusal to mate) has been noted in field studies of several other mammalian species²¹⁻²³. □

Received 4 June; accepted 11 September 1990.

- Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (Murray, London, 1871).
- Parker, G. A. *Biol. Rev. Camb. Philos. Soc.* **45**, 525-567 (1970).
- Trivers, R. in *Sexual Selection and the Descent of Man 1871-1971* (ed. Campbell, B.) 136-179 (Aldine-Atherton, Chicago, 1972).
- Dewsbury, D. A. *Am. Nat.* **119**, 601-610 (1982).
- Rutowski, R. L. *Fla Ent.* **65**, 72-82 (1982).
- Parker, G. A. in *Mate Choice* (ed. Bateson, P.) 141-166 (Cambridge University Press, 1983).
- Forsberg, J. *Oikos* **49**, 46-54 (1987).
- Gwynne, D. T. *Science* **213**, 779-780 (1981).
- Thornhill, R. & Alcock, J. *The Evolution of Insect Mating Systems* (Harvard University Press, 1983).
- Côté, I. M. & Hunte, W. *Anim. Behav.* **38**, 78-88 (1989).
- Parker, G. A. in *Behavioural Ecology: an Evolutionary Approach* (ed. Krebs, J. R. & Davies, N. B.) 214-244 (Blackwell, Oxford, 1978).
- Schwagmeyer, P. L. *Anim. Behav.* **34**, 297-298 (1986).
- Schwagmeyer, P. L. & Wootner, S. J. *Behav. Ecol. Sociobiol.* **17**, 291-296 (1985).
- Foltz, D. W. & Schwagmeyer, P. L. *Am. Nat.* **133**, 257-265 (1989).
- Foster, M. A. *Am. J. Anat.* **54**, 487-511 (1934).
- Schwagmeyer, P. L. & Foltz, D. W. *Anim. Behav.* **39**, 156-162 (1990).

- Schwagmeyer, P. L. & Parker, G. A. *Anim. Behav.* **35**, 1015-1025 (1987).
- MacArthur, R. H. & Pianka, E. R. *Am. Nat.* **100**, 603-609 (1966).
- Stephens, D. W. *Anim. Behav.* **33**, 667-669 (1985).
- Beach, F. A. *Hormones and Behav.* **7**, 105-138 (1976).
- Farentinos, R. C. *Anim. Behav.* **20**, 316-326 (1972).
- Packer, C. & Pusey, A. E. *Am. Nat.* **121**, 716-728 (1983).
- Hogg, J. T. *Behav. Ecol. Sociobiol.* **22**, 49-59 (1988).
- McNamara, J. M. & Houston, A. I. *Anim. Behav.* **35**, 1084-1099 (1987).

ACKNOWLEDGEMENTS. We thank S. Forbes and D. Mock for their comments on the manuscript, T. Lamey for computer analyses, and D. Mock, especially, for assistance with the work. The research was supported by the National Science Foundation, the University of Oklahoma Research Council, and the Natural Environment Research Council.

Molecular and phenotypic variation in the *achaete-scute* region of *Drosophila melanogaster*

Trudy F. C. Mackay* & C. H. Langley†

* Department of Genetics, Box 7614, North Carolina State University, Raleigh, North Carolina 27695-7614, USA

† Center for Population Biology and Department of Genetics, University of California, Davis, California 95616, USA

VARIATION in quantitative characters underlies much adaptive evolution and provides the basis for selective improvement of domestic species, yet the genetic nature of quantitative variation is poorly understood¹. Many loci affecting quantitative traits have been identified by the segregation of mutant alleles with major qualitative effects^{2,3}. These alleles may represent an extreme of a continuum of allelic effects, and most quantitative variation could result from the segregation of alleles with subtle effects at loci identified by alleles with major effects⁴⁻⁶. The *achaete-scute* complex in *Drosophila melanogaster* plays a central part in bristle development^{7,8} and has been characterized at the molecular level^{9,10}. The hypothesis that naturally occurring quantitative variation in bristle number could be associated with wild-type alleles of *achaete-scute* was tested by correlating phenotypic variation in bristle number with molecular variation in restriction maps in this region among chromosomes extracted from natural populations. DNA insertion variation in the *achaete-scute* region was found to be strongly associated with variation in bristle number.

The restriction map variation of a 106-kilobase region including the *achaete-scute* complex has been quantified among 64 isogenic X-chromosome lines extracted from three natural *D. melanogaster* populations¹¹. Thirty-six of the X chromosomes from two of the populations (Raleigh and Texas) studied in this survey were still available (from C. C. Laurie, Duke University). Variation among these chromosomes for abdominal and sternopleural bristle number was assessed. The autosomes of all X chromosome lines were isogenic (see ref. 12 for details), so variation in bristle score among the lines is due to variation at X chromosome loci. There are 17 restriction map polymorphisms in the *achaete-scute* region among the 36 lines, with 4 polymorphic restriction sites, 11 insertions and 2 deletions. Of these polymorphisms, only 6 have frequencies greater than 0.10. Because of such low heterozygosities, and the large amount of phenotypic variation for the bristle traits, we did not attempt to associate presence or absence of individual molecular variants with variation in bristle score. As known bristle mutants are associated with lesions in the *achaete-scute* region^{9,13}, we tested whether the insertions in the naturally occurring *achaete-scute* alleles were associated with phenotypic variation in abdominal and sternopleural bristle number. Of the 13 haplotypes represented in the sample, 9 have at least one insertion. The insertion haplotypes are represented by 16 of 36 lines.

The mean rank order bristle scores of the 36 X-chromosome lines and their insertion haplotypes are shown in Fig. 1. Averaged over both populations, lines with insertions have 1.62 fewer sternopleural bristles and 1.18 fewer abdominal bristles than

lines without insertions. This difference was highly significant when tested by analysis of variance (Table 1). In terms of additive genetic (σ_a) and phenotypic (σ_p) standard deviation units, the reduction in bristle number for haplotypes containing insertions compared with insertion-free haplotypes is $1.62\sigma_a$ and $0.73\sigma_p$ for sternopleural bristles, and $1.53\sigma_a$ and $0.58\sigma_p$ for abdominal bristles.

The significant difference in overall female and male score of 0.74 sternopleural and 2.89 abdominal bristles is typical of sexual dimorphism for this species. The sex dimorphism in bristle number varies significantly among lines for both traits, suggesting that genes on the X chromosome affecting bristle number have differential effects on the two sexes. Interestingly, the presence or absence of insertions in *achaete-scute* has a significant effect on the sex dimorphism of abdominal bristle number (see Table 1; sex by insertion interaction).

To our knowledge, this is the first evidence that insertions in natural populations are associated with quantitative morphological variation, although there is evidence that new *P*-element insertions cause variation for bristle number and other quantitative traits¹⁴⁻¹⁶. The fraction of the X-chromosomal variation for bristle number attributable to insertional variation at *achaete-scute* is surprisingly large. If we assume all insertion haplotypes have similar additive effects in reducing bristle score, and the insertion haplotypes are in Hardy-Weinberg equilibrium, then the variance contributed by insertions at *achaete-scute* in a random-mating population is $a^2q(1-q)/2$, where a is the difference in mean bristle number between homozygous insertion and non-insertion haplotypes, and q is the frequency of insertion haplotypes. The additive variance due to insertions at *achaete-scute* is thus 0.323 for sternopleural and 0.172 for abdominal bristle number, respectively; and the proportion of

TABLE 1 Analysis of variance of sternopleural and abdominal bristle number

Source of variation	d.f.	Sternopleural bristles		Abdominal bristles	
		Mean square	F	Mean square	F
Sex	1	195.81	24.07*	3,013.12	351.18*
Pop	1	264.19	2.92 ^{ns}	11.13	0.20 ^{ns}
Ins	1	856.53	9.46†	484.35	8.66†
Pop × Ins	1	13.82	0.15 ^{ns}	2.22	0.04 ^{ns}
Sex × Pop	1	0.39	0.05 ^{ns}	0.10	0.01 ^{ns}
Sex × Ins	1	0.10	0.01 ^{ns}	62.99	7.34‡
Sex × Pop × Ins	1	0.14	0.02 ^{ns}	3.01	0.35 ^{ns}
Line (Pop × Ins)	32	90.51	13.76*	55.91	12.14*
Sex × Line (Pop × Ins)	32	8.14	1.81‡	8.58	1.74‡
Rep (Line)	36	6.58	1.46 ^{ns}	4.61	0.93 ^{ns}
Sex × Rep (Line)	36	4.50	1.30 ^{ns}	4.94	1.54‡
Error	1,296	3.47		3.22	

Variance was partitioned into sources attributable to main independent fixed effects due to sex (Sex), population (Pop), insertion categories (Ins) and their two- and three-way interactions; the random effect of chromosome line (Line) nested within population by insertion category; the random effect of replicate vial (Rep) nested within chromosome line; sex by line within population by insertion class and sex by replicate within line interactions; and error. The analysis was performed using SAS variance components estimation procedure. The variance component among homozygous X-chromosome lines (within population and insertion categories) estimates twice the X-chromosomal genetic variance for bristle score of a non-inbred population, assuming complete additivity. The estimate of additive genetic variance due to loci on the X chromosome from this analysis is thus 1.004 for sternopleural and 0.596 for abdominal bristle number. The phenotypic variance is estimated from the sum of all variance components, and is 4.859 for sternopleural and 4.168 for abdominal bristle number. The square root of the additive genetic (phenotypic) variance gives the additive genetic (phenotypic) standard deviation units used in the text to scale the effect of insertion haplotypes on bristle score.

* $P < 0.001$; † $0.001 < P < 0.01$; ‡ $0.01 < P < 0.05$; ns, $P > 0.05$.

the total X-chromosome genetic variance due to effects of inserts at *achaete-scute* is 0.243 for sternopleural bristles and 0.224 for abdominal bristles. As the X chromosome is roughly one-fifth of the genome, about 5% of the total genetic variance of abdominal and sternopleural bristle number might be caused by the effects of insertions at this locus.

This analysis assumes that *achaete-scute* is in linkage equilibrium with other X-chromosome loci. Because *achaete-scute* is located at the tip of the X chromosome, in an area of reduced recombination¹¹, it is possible that the observed effect is caused by linkage disequilibrium between *achaete-scute* and linked loci. But the restriction of recombination does not extend beyond the *zeste-white* region, for which pairs of closely linked polymorphic restriction sites show little linkage disequilibrium¹¹, so any linkage disequilibrium involves only loci in a small region of the X chromosome immediately adjacent to *achaete-scute*.

We have shown that naturally occurring molecular variation at a major bristle locus, the *achaete-scute* complex, can be associated with nearly 5% of the total quantitative genetic variation for bristle number. It is likely that a small number of loci contribute most of the genetic variation for any quantitative trait, with more numerous loci contributing less^{17,18}. This analysis suggests candidates for the loci with most effect on a trait may be loci identified by major qualitative mutations affecting fundamental developmental processes, and that the approach taken here with *achaete-scute* may be applicable in general to the study of quantitative genetic variation. The nature of the restriction map variation associated with the phenotypic variation is intriguing. It is important to determine whether naturally occurring insertions generally have detectable phenotypic effects, and if so, if they commonly reduce trait values. The *achaete-scute* complex may be peculiar because of its large regulatory regions¹³, its interactions with genes involved in sex determination¹⁹, and because of its location at the tip of the X chromosome¹¹. □

Received 14 June; accepted 20 August 1990.

1. Falconer, D. S. *Introduction to Quantitative Genetics* (Longman, Harlow, 1989).
2. Lindsley, D. L. & Grell, E. H. *Genetic Variations of Drosophila melanogaster* (Carnegie Inst. Washington, 1968).

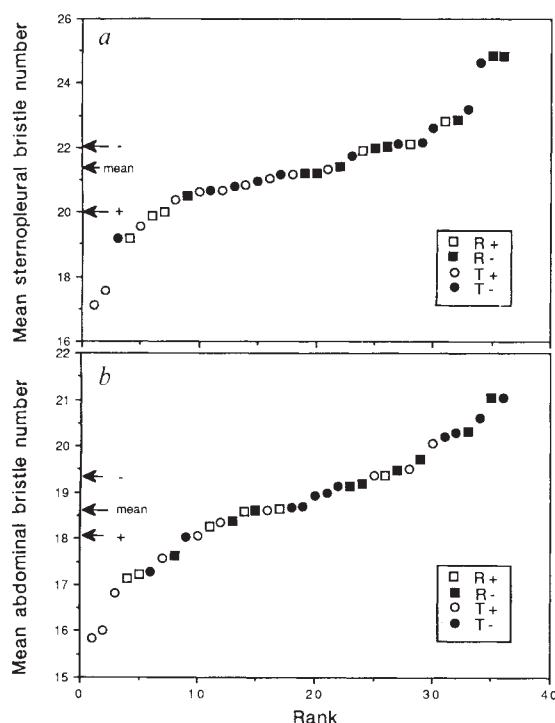


FIG. 1 Rank order of mean sternopleural (a) and abdominal (b) bristle scores from isogenic X chromosome lines extracted from the Texas (T) and Raleigh (R) natural populations¹². The total number of sternopleural bristles on the right and left sternopleural plates, and the number of abdominal bristles on the most posterior abdominal sternite, was recorded on a sample of 10 males and 10 females from each of 2 replicate vials per line. Each line mean is the average of these 40 measurements. The presence of an insertion in the *yellow-achaete-scute* region¹¹ is indicated in the box by +, and no insertion in the region by -. The arrows on the ordinate of the plots indicate the mean bristle scores of all lines (mean), of lines without insertions (-) and of lines with insertions (+).

3. Mackay, T. F. C. in *Evolution and Animal Breeding* (eds Hill, W. G. & Mackay, T. F. C.) 113-119 (C.A.B. International, Wallingford, 1989).
4. Thompson, J. N. Jr *Nature* **258**, 665-668 (1975).
5. Robertson, D. S. *J. theor. Biol.* **117**, 1-10 (1985).
6. Mackay, T. F. C. *Genetics* **111**, 351-374 (1985).
7. Ghysen, A. & Dambly-Chaudière, C. *Trends Genet.* **5**, 251-255 (1989).
8. Ghysen, A. & Dambly-Chaudière, C. *Genes Dev.* **2**, 495-501 (1988).
9. Campuzano, S. *et al. Cell* **40**, 327-338 (1985).
10. Villares, R. & Cabrera, C. V. *Cell* **50**, 415-424 (1987).
11. Aguadé, M., Miyashita, N. & Langley, C. H. *Genetics* **122**, 607-615 (1989).
12. Miyashita, N., Laurie-Ahlberg, C. C., Wilton, A. N. & Emigh, T. H. *Genetics* **113**, 321-335 (1986).
13. Ruiz-Gómez, M. & Modolell, J. *Genes Dev.* **1**, 1238-1246 (1987).
14. Mackay, T. F. C. *Genet. Res.* **49**, 225-233 (1987).
15. Mackay, T. F. C. *Genet. Res.* **48**, 77-87 (1986).
16. Lai, C. & Mackay, T. F. C. *Genetics* **124**, 627-636 (1990).
17. Robertson, A. in *Heritage from Mendel* (ed. Brink, A.) 265-280 (Univ. Wisconsin Press, Madison, 1967).
18. Shrimpton, A. E. & Robertson, A. *Genetics* **118**, 445-459 (1988).
19. Hartley, D. & White, R. *Trends Genet.* **6**, 199-201 (1990).

ACKNOWLEDGEMENTS. We thank R. Lyman for statistical analysis of the data, and C. C. Cockerham, W. G. Hill, E. Holmes, M. Turelli and B. Walsh for comments on the manuscript. This work was supported by the NIH.

Stimulation of the phosphatidylinositol pathway can induce T-cell activation

Dev M. Desai, Marianne E. Newton, Theresa Kadlecsek & Arthur Weiss

Howard Hughes Medical Institute, Departments of Medicine and of Microbiology and Immunology, University of California at San Francisco, San Francisco, California 94143, USA

THE T-cell antigen receptor (TCR) regulates two signal transduction pathways: the phosphatidylinositol (PtdIns)¹ and tyrosine kinase pathways². Stimulation of T cells with antigen or anti-TCR monoclonal antibodies induces an increase in inositol phosphates and diacylglycerol^{1,3}, the second messengers responsible for the mobilization of cytoplasmic free calcium and activation of protein kinase C⁴. The TCR also activates a tyrosine kinase that is not intrinsic to the TCR². The relationship between these two signal transduction pathways and their contribution to later T-cell responses is unclear. Studies using variants of a murine hybridoma suggested that the PtdIns pathway might not be necessary for or be involved in regulating interleukin-2 (IL-2) production⁵. To address the relationship between later T-cell responses and the early biochemical signals, we investigated the ability of a heterologous receptor with defined signal transduction function to induce T-cell activation. The human muscarinic subtype-1 receptor (HM1)⁶, which elicits PtdIns metabolism in neuronal cells through a G protein-coupled mechanism⁷, also functionally activates this pathway when expressed in the T-cell line Jurkat-derived host, J-HM1-2.2 (ref. 8). We show here that stimulation of HM1 alone induced IL-2 production and IL-2 receptor α chain expression. HM1 does not induce the tyrosine kinase pathway, suggesting that this pathway does not directly influence later T cell-activation responses. Instead, our studies indicate that activation of the PtdIns pathway is probably sufficient to induce later T-cell responses.

Comparison of the time-dependent appearance of inositol phosphates revealed that Jurkat T cells responded well to stimulation by C305, an agonist anti-TCR monoclonal antibody, but were unresponsive to the muscarinic receptor agonist carbachol (Fig. 1a), consistent with its failure to express muscarinic binding sites⁸. J-HM1-2.2, which expresses both receptors, increased all three inositol phosphate fractions in response to either C305 or carbachol (Fig. 1b). When HM1 was stimulated, however, the total increase of inositol phosphates was substantially greater and their temporal accumulation was more prolonged than in the TCR-induced response.

As comparable numbers of TCR and HM1 are expressed on J-HM1-2.2 (ref. 8), the differences in observed responses must

be attributable to other differences between these receptors. The TCR is known to be internalized following exposure to soluble anti-TCR monoclonal antibody^{9,10}. Unlike the TCR, whose expression decreases substantially during stimulation, expression of HM1 is not diminished by carbachol stimulation (Fig. 1b). Hence, failure of HM1 to internalize may contribute to the larger and more sustained increases in inositol phosphates following ligand binding. In contrast, the more transient inositol phosphate increases induced by the TCR are likely to be attributable, in part, to its rapid internalization after binding soluble

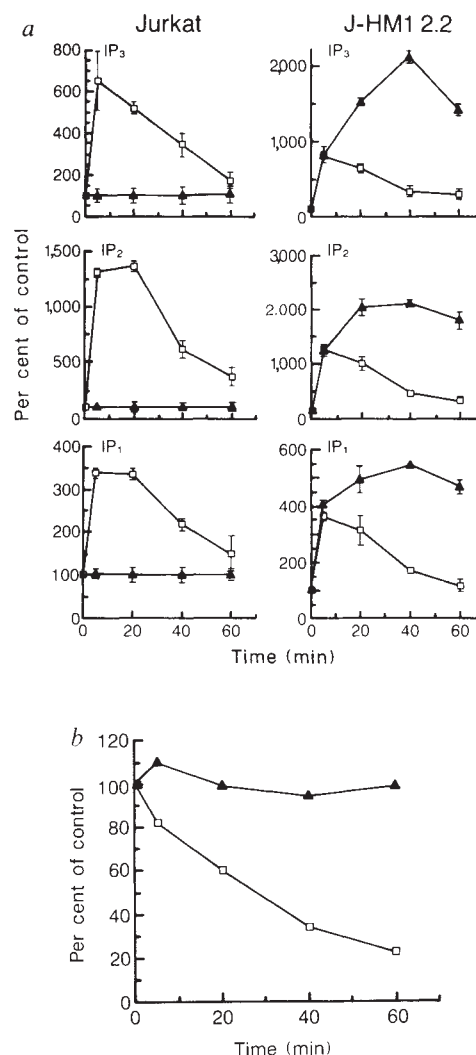


FIG. 1 a, Jurkat or J-HM1-2.2 cells were stimulated with anti-TCR monoclonal antibody (mAb)C305 (□) or carbachol (▲) and the indicated inositol fractions were assayed at the times indicated. IP₃, inositol-1,4,5-trisphosphate; IP₂, inositol-1,4-bisphosphate; IP₁, inositol-1-monophosphate. b, J-HM1-2.2 cells were stimulated with saturating amounts of either C305 (□) or carbachol (▲) as in a in medium at 37 °C. At various times, cells were removed and the relative levels of the TCR or HM1 receptors remaining on the cell surface were determined.

METHODS. a, Inositol phosphate levels in Jurkat and J-HM1-2.2 following stimulation were determined as described⁸. Cells were stimulated with either C305 (1/1,000 dilution of ascitic fluid, a saturating concentration) or carbachol (500 μ M, a saturating concentration). Results are expressed as the mean % of control (unstimulated cells) $\pm$ s.e.m. b, Relative levels of the TCR or HM1 receptors remaining on the cell surface after incubation with anti-TCR (C305) or carbachol were determined by immunofluorescence with fluorescein-conjugated anti-Leu 4 mAb (anti-CD3) and flow cytometry or by saturable binding studies with the muscarinic antagonist [³H]quinuclidyl benzilate (QNB), respectively⁸. Results are expressed as the mean % of the level of relevant receptor expression on unstimulated cells from three separate experiments.

TABLE 1 Stimulation of the HM1 receptor expressed on J-HM1-2.2 induces IL-2 production

Stimulus	Cell	
	Jurkat	J-HM1-2.2
	IL-2 (units ml ⁻¹)	
Medium	<1	<1
PMA	<1	<1
Ionomycin	<1	<1
C305 (anti-TCR)	<1	<1
Carbachol	<1	33
Ionomycin + PMA	320	>2,840
C305 + PMA	84	155
Carbachol + PMA	<1	1,279

Jurkat or J-HM1-2.2 cells were stimulated as indicated for 24 h. IL-2 content in the resulting cell-free supernatants was determined using the CTLL 2.20 IL-2 dependent cell line in the MTT bioassay as described³².

anti-TCR monoclonals. This may represent a protective mechanism, as T cells normally respond to antigens that are associated with cell surfaces rather than in soluble form, and generally respond better to anti-TCR monoclonals that have been immobilized^{11,12}. This is consistent with the more prolonged translocation of protein kinase C (PKC)¹³ and inositol phosphate generation (data not shown) in response to immobilized compared with soluble anti-CD3 monoclonal antibodies.

T-cell activation is associated with induction of IL-2 receptor α chain (gp55)^{2,14} and CD69 (ref. 15). The ability of HM1 to induce these responses was compared with the TCR-induced expression of these molecules (Fig. 2). Incubation of Jurkat or J-HM1-2.2 with phorbol myristyl acetate (PMA), which activates PKC, or with the anti-TCR monoclonal antibody C305 resulted in the expression of the α chain of the IL-2 receptor. Stimulation of J-HM1-2.2, but not Jurkat, with carbachol induced IL-2 receptor α chain expression on a large proportion of cells. Heterogeneity in the carbachol-induced expression of the IL-2 receptor α chain was consistently observed and reflects clonal heterogeneity in the response as all J-HM1-2.2 cells increase cytoplasmic free calcium concentration ($[Ca^{2+}]_i$) in response to carbachol (data not shown). Comparable results were observed when the expression of CD69 was assessed (data not shown). As induction of the IL-2 receptor α chain involves activation of α -chain gene transcription¹⁴, signal transduction by HM1 can influence cellular responses that depend on transcription.

The requirements for IL-2 production are more stringent than those for IL-2 receptor and CD69 expression³. As in previous studies¹⁶, anti-TCR monoclonal antibody or calcium ionophore induced Jurkat or J-HM1-2.2 cells to produce substantial amounts of IL-2 only in the presence of PMA (Table 1). However, J-HM1-2.2 produced a substantial but submaximal amount of IL-2 when stimulated with carbachol alone. This

response was potentiated by the addition of PMA. The ability of carbachol alone to induce IL-2 production may relate to the greater or more sustained second messenger generation elicited by HM1 stimulation. Sustained signal transduction by the TCR is required for cellular commitment to IL-2 production¹⁷. As the responses initiated by the TCR or HM1 receptor are potentiated by PMA, this suggests that stimulation of neither receptor maximally activates PKC, consistent with previous findings for the TCR¹³.

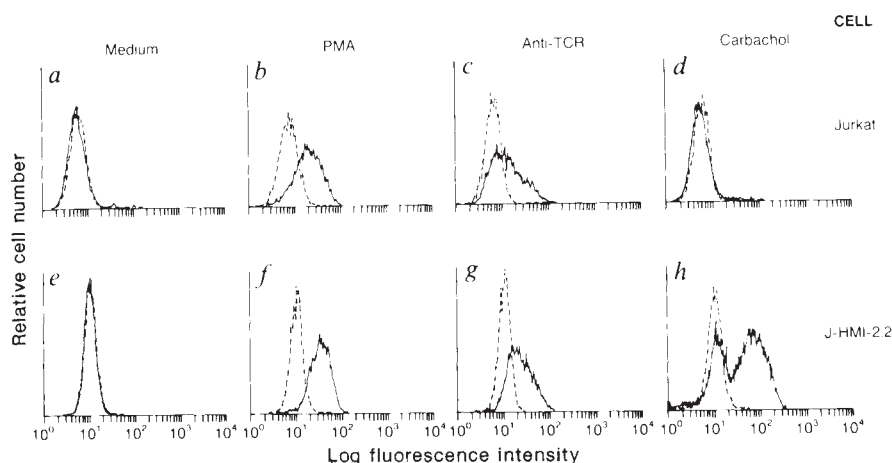
IL-2 production can also be induced by stimulation of other T-cell surface receptors including DC2 and Thyl^{18,19}, but requires expression of a functional TCR²⁰. J-HM1-2.1, a Jurkat-derived HM1 transfectant, fails to express the TCR but HM1 still activates the PtdIns pathway⁸. J-HM1-2.1, like the TCR⁺, J-HM1-2.2 could also produce IL-2 and express IL-2 receptor α chain when stimulated with carbachol (data not shown). Thus, HM1 can function independently of the TCR to activate the PtdIns signal transduction pathway and induce later responses associated with T-cell activation.

Cell-surface molecules other than the TCR contribute to T-cell activation. CD28 can deliver a signal which synergizes with TCR agonists in inducing IL-2 production^{21,22}, but its signal transduction mechanism is distinct from that of the TCR^{22,23}. If HM1 activates IL-2 production by way of a signal transduction mechanism also used by the TCR, activation by HM1 should be potentiated by stimulation of CD28. J-HM1-2.2 cells were stimulated with anti-TCR monoclonal antibody plus PMA or carbachol in the presence or absence of PMA; all these stimuli induce IL-2 production (Table 1). Costimulation with anti-CD28 (9.3) but not an anti-HLA (w6/32) resulted in substantial augmentation of IL-2 production induced by the TCR or HM1 (Fig. 3). These results suggest that signals mediated by HM1 that lead to IL-2 production qualitatively mimic those of the TCR.

As the TCR activates a tyrosine kinase signal transduction pathway, we determined whether HM1 also activates the TCR-regulated tyrosine kinase pathway. In western blots of whole cell lysates probed with an anti-phosphotyrosine monoclonal antibody, stimulation of Jurkat and J-HM1-2.2 cells with anti-TCR induced increased tyrosine phosphorylation of several proteins (Fig. 4a). The most prominent and consistent changes after TCR stimulation were induced or increased phosphorylation of pp19–22, pp38, pp42–44, pp80, pp100 and pp135. These changes peaked by 2 min and declined substantially after 15 min of stimulation (data not shown). Variable changes in several phosphoproteins in the relative molecular mass range 52,000–62,000 (M_r 52–62K) were observed. The changes in pp100 and pp135 were best visualized with a rabbit anti-phosphotyrosine serum, although this antiserum was not as sensitive in detecting other substrates. In contrast, the only tyrosine phosphoprotein induced by HM1 during 2 h of stimulation was the pp42–44 doublet. In some experiments, pp42–44 has been constitutively tyrosine-phosphorylated. Ionomycin did not alter tyrosine phos-

FIG. 2 Induction of IL-2 receptor α chain expression in Jurkat (a–d) or J-HM1-2.2 cells (e–h) was assessed following stimulation with medium (a, e); PMA, 50 ng ml⁻¹ (b, f); C305, 1/1,000 of ascitic fluid (anti-TCR; c, g); or carbachol, 500 μ M (d, h) for 24 h.

METHODS. Cells were stained with fluoresceinated conjugates of a control IgG mAb (—) or a mAb reactive with the α chain of the IL-2 receptor (—) (Becton Dickinson). Immunofluorescence intensity of the cells was determined by flow cytometry.



phoproteins (data not shown), but PMA also induced pp42-44. A 42K PKC-dependent tyrosine phosphoprotein has also been observed in response to platelet-derived growth factor receptor stimulation and is believed to represent microtubule-associated protein 2 kinase^{24,25}. Moreover, the requirement for receptor-mediated activation of PKC for the appearance of pp42-44 is supported by studies of a signal transduction mutant of Jurkat, J.CaM2, in which stimulation of the TCR results in activation of the tyrosine kinase pathway but not the PtdIns pathway and pp42-44 is not detected (A.W., manuscript in preparation). Therefore, although TCR stimulation results in the appearance of several new tyrosine phosphoproteins, HM1 stimulation induces only pp42-44, whose appearance can be attributed to the activation of PKC through induction of the PtdIns pathway.

We also investigated the ability of HM1 to induce tyrosine phosphorylation of the only well-characterized substrate of the TCR-activated tyrosine kinase, the ζ chain, in immunoprecipitation studies. Direct TCR occupancy is not required to induce ζ phosphorylation as activating anti-Thy 1 or anti-CD2 monoclonals can induce ζ phosphorylation^{26,27}, although these molecules may interact with CD3 during activation^{2,20}. Figure 4b shows that TCR, but not HM1, stimulation leads to increases in ζ phosphorylation. Only a small proportion of total cellular ζ is phosphorylated following TCR stimulation (compare Fig. 4b and c), consistent with observations in the murine 2B4 hybridoma²⁶.

The TCR and HM1 activate overlapping but not identical signal transduction pathways. They both activate the PtdIns

pathway, but HM1 does not activate the TCR-regulated tyrosine kinase. Recent studies with herbimycin, an inhibitor of Src-like protein kinases²⁸ which inhibits TCR-induced activation of the PtdIns pathway²⁹, support these findings. In J-HM1-2.2, herbimycin inhibits TCR but not HM1 activation of the PtdIns

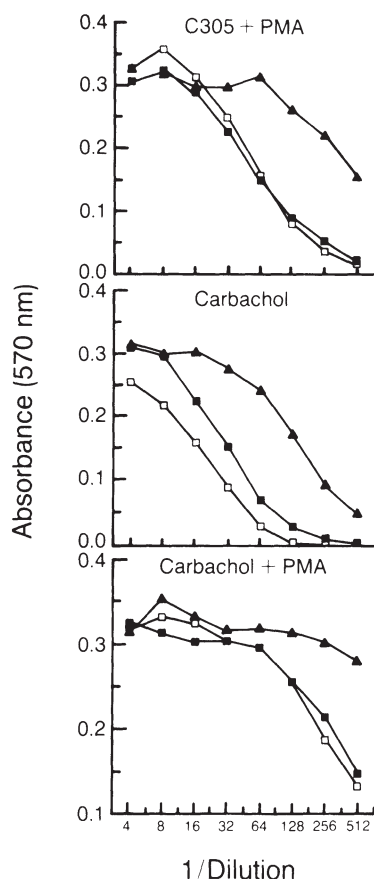


FIG. 3 J-HM1-2.2 cells were stimulated with anti-TCR (C305) plus PMA or with carbachol in the absence or presence of PMA. To the indicated replicate sets of cultures, medium ($\square$) or saturating quantities of mAb reactive with CD28 (mAb 9.3; $\blacktriangle$) or class I HLA (mAb w6/32; $\blacksquare$) were added. Supernatants were assayed for IL-2 biological activity after 24 h of culture using an IL-2 dependent cell line in a colorimetric bioassay³². Cell viability is proportional to the amount of converted dye that absorbs light at 570 nm.

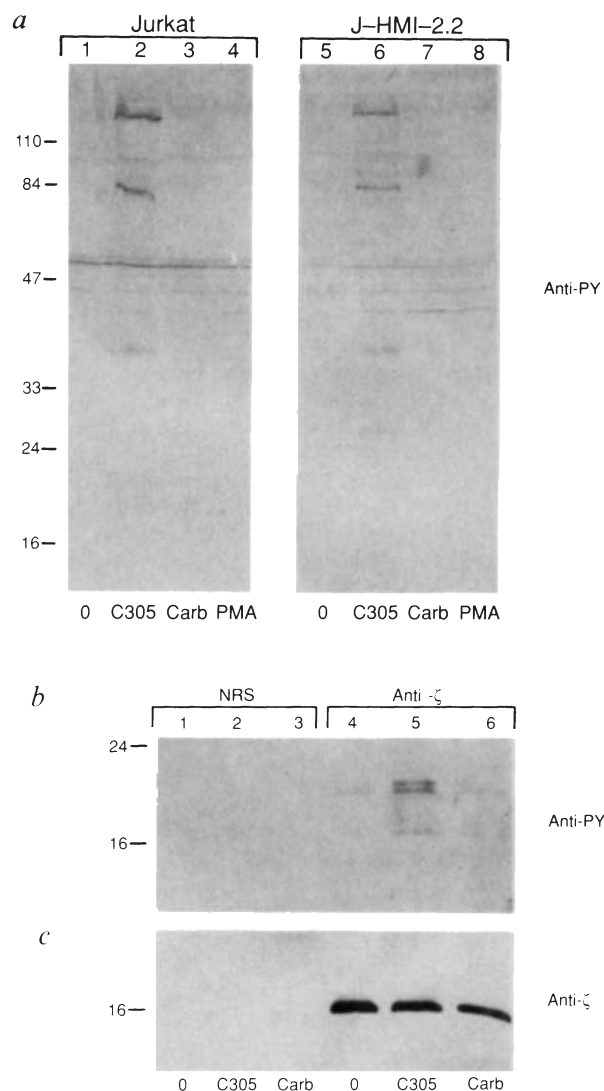


FIG. 4 a, Tyrosine phosphoproteins detected by western blot analysis. Whole cell lysates were prepared from Jurkat (1-4) or J-HM1-2.2 cells (5-8) stimulated for 2 min with medium (1, 5), C305 (2, 6), carbachol (3, 7), or PMA (4, 8). b, c, Western blots of normal rabbit serum (1-3) or rabbit anti- ζ peptide antiserum (4-6) immunoprecipitates probed with an anti-phosphotyrosine mAb (b) or rabbit anti- ζ peptide antiserum (c). Cells were stimulated with medium (1, 4); anti-TCR (C305; 2, 5); or carbachol (3, 6). METHODS. Cells, $5 \times 10^6 \text{ ml}^{-1}$, were stimulated for 2 min at 37°C with the following stimuli in serum-free medium: carbachol, 500 μM , C305 (1/1,000 dilution of ascitic fluid, a saturating concentration), and PMA, 50 ng ml^{-1} . Cells were solubilized in 1% NP-40 in Tris-buffered saline, pH 7.5, containing protease (PMSF, aprotinin, pepstatin and leupeptin) and phosphatase inhibitors (EDTA, sodium fluoride, sodium vanadate and pyrophosphate). For a, solubilized proteins derived from 2×10^6 cells were resolved by SDS-PAGE, and transferred to nitrocellulose. Immunoprecipitates were isolated from lysates of 5×10^7 cells stimulated for 5 min at 37°C in medium containing 10% fetal calf serum. Immunoprecipitates were prepared with normal rabbit serum or rabbit anti- ζ peptide (rabbit 387 anti-peptide 3)²⁶. Proteins containing phosphotyrosine or ζ -chain were identified using mAb 4G10 (ref. 33) or rabbit anti- ζ , respectively, followed by alkaline phosphatase conjugated-goat anti-mouse or rabbit IgG. The specificity of 4G10 for phosphotyrosine-containing proteins on these blots was confirmed in competition studies using free phosphotyrosine, phosphoserine and phosphothreonine. Molecular mass standards (in thousands) are shown at left.

pathway and IL-2 production (M. Graber and A.W., manuscript in preparation). Moreover, unlike the TCR, HM1 activation of the PtdIns pathway is not dependent on the tyrosine phosphatase CD45 (ref. 30).

These results with HM1 are therefore an example of receptor-mediated PtdIns pathway activation in the absence of tyrosine kinase activation in T cells. Thus, HM1 activation is a valuable model for testing the role of these two signal transduction pathways in later T-cell activation responses. Activation of the PtdIns pathway through HM1 extends previous observations made with calcium ionophores and phorbol esters¹⁶, as changes in $[Ca^{2+}]_i$ and PKC activity induced by these pharmacological agents may be regulated through distinct mechanisms and have other unidentified effects. Whether the TCR and HM1 have other transmembrane signalling systems in common is not known, but it is clear that by using a signalling pathway shared by the TCR, HM1 can induce several later T-cell responses. Although there may be other signal transduction pathways leading to T-cell activation, the weight of evidence would suggest that the ability of these two structurally distinct receptors to activate the inositol phospholipid pathway is sufficient to explain their effects on T-cell activation responses. These studies do not exclude the possibility that the tyrosine kinase or other TCR-induced signal transduction pathways can regulate IL-2 production, as demonstrated in the murine 2B4 hybridoma model⁵. Moreover, the tyrosine kinase pathway may have a critical role in regulating the PtdIns pathway, perhaps by activating phospholipase C, as has been suggested with other nonlymphoid tyrosine kinase receptors³¹. Finally, these studies demonstrate that ectopic expression of receptors that couple to the PtdIns pathway could lead to antigen-independent activation of T cells which might lead to pathological autoimmune phenomena. □

HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside the CD4-binding domain

William A. O'Brien*, Yoshio Koyanagi†, Ali Namazie‡, Jia-Qi Zhao‡, Amadou Diagne‡, Ken Idler§, Jerome A. Zack‡ & Irvin S. Y. Chen‡

* Division of Infectious Diseases, Department of Medicine, West Los Angeles Veteran's Administration Medical Center, UCLA School of Medicine, Los Angeles, California 90073, USA

† Department of Virology and Parasitology, Yamaguchi University School of Medicine, Ube, Yamaguchi 755, Japan

‡ Departments of Microbiology & Immunology and Medicine, Jonsson Comprehensive Cancer Center, UCLA School of Medicine, Los Angeles, California 90024-1678, USA

§ Abbott Laboratories, Department 93D, AP9A/3, 1400 Sheridan Avenue, Abbott Park, Illinois 60064, USA

CELLS of the mononuclear phagocyte system are the predominant cell producing HIV-1 in most tissues, including the central nervous system (CNS)¹⁻⁴, spinal cord⁵, lung⁶ and skin^{7,8}; infection is associated with dementia^{9,10}, neuropathy¹¹⁻¹³, pneumonitis^{9,14} and dermatitis^{15,16} respectively. Different HIV-1 isolates vary markedly in their ability to infect mononuclear phagocytes productively^{17,18}. Here we describe molecular clones of a CNS-derived isolate, HIV-1_{JR-FL}, which can replicate efficiently in mononuclear phagocytes. Analysis by polymerase chain reaction of early events after infection indicates that the early phase of viral replication before reverse transcription determines tropism. Genetic mapping of the macrophage-tropic phenotype by construction of recombinant viruses indicates that mononuclear phagocyte infectivity can be determined by a 157-amino-acid region of the gp120 glycoprotein of HIV-1_{JR-FL}. Significantly, this region is upstream from the previously defined CD4-binding domain¹⁹. We propose that at least one determinant for mononuclear phagocyte tropism involves target cell interactions with regions of gp120 distinct from the CD4-binding domain.

HIV-1_{JR-FL} was recovered from frontal lobe brain tissue of a patient (J.R.) who died with AIDS and severe AIDS encephalopathy¹⁸. The brain showed extensive leukoencephalopathy with numerous multinucleated giant cells in sections taken at autopsy. In an effort to isolate virus and subsequently molecular clones that would closely represent a primary clinical isolate, we obtained virus by short-term cocultivation of frontal lobe brain tissue with lectin-stimulated human peripheral blood lymphocytes (PBL). The virus recovered replicated efficiently in PBL and mononuclear phagocytes. In contrast to many other HIV-1 isolates, HIV-1_{JR-FL} was not passaged in tumour cell lines and in fact does not replicate in Jurkat, U937 or HUT 78 cells (our unpublished data). Before investigating HIV-1_{JR-FL} replication in mononuclear phagocytes, we compared the ability of several cloned HIV-1 strains to infect primary blood mononuclear phagocytes productively. All strains tested replicate efficiently in phytohaemagglutinin-activated PBL. Infection of mononuclear phagocytes with HIV-1_{JR-CSF}¹⁸ or HIV-1_{SF2}²⁰ is less efficient than with HIV-1_{JR-FL}, but usually results in virus production (data not shown). Virus replication following infection with HIV-1_{NL4-3}²¹, however, was least efficient. There is usually no detectable virus following HIV-1_{NL4-3} infection of mononuclear phagocytes, although low levels of p24 could occasionally be detected, but these were at least 10-fold lower than after HIV-1_{JR-FL} infection. We therefore chose to compare HIV-1_{NL4-3} with HIV-1_{JR-FL}.

Subgenomic molecular clones of HIV-1_{JR-FL}, which together comprise the entire genome, were derived from viral DNA

Received 22 February; accepted 16 August 1990.

- Imboden, J. B. & Stobo, J. D. *J. exp. Med.* **161**, 446-456 (1985).
- Samelson, L. E., Patel, M. D., Weissman, A. M., Harford, J. B. & Klausner, R. D. *Cell* **46**, 1083-1090 (1986).
- Weiss, A. & Imboden, J. B. *Adv. Immun.* **41**, 1-38 (1987).
- Berridge, M. J. *A. Rev. Biochem.* **56**, 159-193 (1987).
- Sussman, J. J. *et al. Nature* **334**, 625-628 (1988).
- Nathanson, N. M. *A. Rev. Neurosci.* **10**, 195-236 (1987).
- Peralta, E. G. *et al. EMBO J.* **6**, 3923-3929 (1987).
- Goldsmith, M. A., Desai, D. M., Schultz, T. & Weiss, A. *J. Biol. Chem.* **264**, 17190-17197 (1989).
- Meuer, S. C. *et al. J. exp. Med.* **157**, 705-719 (1983).
- Minami, Y., Samelson, L. E. & Klausner, R. D. *J. Biol. Chem.* **262**, 13342-13347 (1987).
- Ledbetter, J. A. *et al. J. Immun.* **136**, 3945-3952 (1986).
- Tax, W. J. M., Hermes, F. F. M., Willems, R. W., Capel, P. J. A. & Koene, R. A. P. *J. Immun.* **133**, 1185-1189 (1984).
- Manger, B., Weiss, A., Imboden, J., Laing, T. & Stobo, J. *J. Immun.* **139**, 395-407 (1987).
- Leonard, W. J., Kronke, M., Pepper, N. J., Depper, J. M. & Green, W. C. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6281-6285 (1985).
- Testi, R., Phillips, J. H. & Lanier, L. L. *J. Immun.* **142**, 1854-1860 (1989).
- Weiss, A., Imboden, J., Shoback, D. & Stobo, J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4169-4173 (1984).
- Weiss, A., Shields, R., Newton, M., Manger, B. & Imboden, J. *J. Immun.* **304**, 445-447 (1987).
- Meuer, S. C. *et al. Cell* **36**, 897-906 (1984).
- Gunter, K. C. *et al. Nature* **326**, 505-507 (1987).
- Bockenstedt, L. K. *et al. J. Immun.* **141**, 1904-1911 (1988).
- Ledbetter, J. A. *et al. J. Immun.* **135**, 2331-2336 (1985).
- Weiss, A., Manger, B. & Imboden, J. *J. Immun.* **137**, 819-825 (1986).
- June, C. H., Ledbetter, J. A., Gillespie, M. M., Lindsten, T. & Thompson, C. B. *Molec. Cell Biol.* **7**, 4472-4481 (1987).
- Kazlauskas, A. & Cooper, J. A. *J. Cell Biol.* **106**, 1395-1402 (1988).
- Rossomando, A. J., Payne, M., Weber, M. J. & Sturgill, T. W. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6940-6943 (1989).
- Orloff, D. G., Frank, S. J., Robey, F. A., Weissman, A. M. & Klausner, R. D. *J. Biol. Chem.* **264**, 14812-14817 (1989).
- Monostori, E., Desai, D., Brown, M. H., Cantrell, D. A. & Crumpton, M. J. *J. Immun.* **144**, 1010-1014 (1990).
- Uehara, Y., Fukazawa, H., Murakami, Y. & Mizuno, S. *Biochem. biophys. Res. Commun.* **163**, 803-809 (1989).
- June, C. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Koretzky, G. A., Picus, J., Thomas, M. L. & Weiss, A. *Nature* **346**, 66-68 (1990).
- Wahl, M. I., Nishibe, S., Suh, P.-G., Rhee, S. G. & Carpenter, G. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1568-1572 (1989).
- Mosmann, T. *J. Immunol. Meth.* **65**, 55-63 (1983).
- Morrison, D. K. *et al. Cell* **58**, 649-657 (1989).

ACKNOWLEDGEMENTS. We thank J. Hansen for monoclonal antibody 9.3, D. Morrison and T. Roberts for 4G10, and L. Samelson, D. Orloff and R. Klausner for the rabbit anti- ζ peptide antiserum used in these studies. We also thank J. Fraser and G. Koretzky for advice and for critical review of this manuscript. This work was supported in part by the NIH. D.M.D. is supported by the NIH Medical Scientist Training Program.

TABLE 1 Replication of parental and recombinant HIV-1 strains in peripheral blood mononuclear phagocytes

Study 1	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	PBL
HIV-1 _{JR-FL}							
Day 14	>40	110	5.4	18	30	1.3	158
Day 21	44	151	>20	13	134	2.4	
HIV-1 _{NF-Sst}							
Day 14	>40	94	NT	>20	>20	0.81	167
Day 21	3.1	>40		12	>20	0.36	
HIV-1 _{NF-Dra}							
Day 14	31	21	14	15	12	0.10	120
Day 21	19	110	>20	3.7	33	0.10	
HIV-1 _{NF-Stu}							
Day 14	>40	NT	11	>20	NT	NT	110
Day 21	>40		>20	5.1			
HIV-1 _{NFN-SX}							
Day 14	>40	NT	NT	>40	NT	NT	160
Day 21	22			15			
HIV-1 _{NF-Xho}							
Day 14	0.18	NT	NT	1.8	NT	NT	110
Day 21	0.42			1.5			
HIV-1 _{NL4-3}							
Day 14	<0.05	<0.05	0.11	1.2	<0.05	<0.05	265
Day 21	<0.05	0.17	0.14	<0.05	0.63	<0.05	
Study 2		<i>g</i>	<i>h</i>	<i>i</i>	<i>j</i>	<i>k</i>	PBL
HIV-1 _{NFN-SX}							
Day 14		47	243	139	0.80	0.57	177
Day 21		43	132	366	0.97	1.2	
HIV-1 _{NFN-SM}							
Day 14		17	19	11	0.47	0.09	151
Day 21		10	15	23	0.68	0.14	
HIV-1 _{NFN-MX}							
Day 14		1.6	13	3.5	<0.05	<0.05	98
Day 21		NT	2.8	0.25	<0.05	<0.05	
HIV-1 _{NL4-3}							
Day 14		1.5	12	1.2	<0.05	<0.05	204
Day 21		1.2	5.8	1.4	0.07	<0.05	

Experiments *a-k* represent infection of mononuclear phagocytes with the indicated virus strains assayed on days 14 or 21. PBL represents infection of peripheral blood mononuclear cells assayed on day 9 (study 1) or day 7 (study 2). Infections in study 1 (*a* to *f*) and study 2 (*g* to *k*) were from independent experiments using mononuclear phagocytes from different donors. HIV-1 p24 values (ng ml⁻¹) are shown.

Virus was recovered by transfection of 5–10 µg of ligated proviral DNA fragments (study 1; see Fig. 1) along with 20 µg pUC 18, or by transfection of 25 µg plasmid from recombinant infectious clones (study 2) into 729 cells (5 × 10⁶) by electroporation³⁷. Experiments *a, b, c, d, e, f, g, h, i, j, k* were performed using virus stocks from separate transfections respectively. Transfected 729 cells were irradiated after 3 days with 10,000 rads, and cocultured along with 729-cell culture supernatants with 2 × 10⁷ phytohaemagglutinin-activated PBL in the presence of 10 µg ml⁻¹ polybrene. Cell-free HIV-1_{JR-FL} supernatant (500 ng) was used to infect 2 × 10⁷ PBL in parallel. In study 1, cultures were fed with autologous phytohaemagglutinin-activated PBL (2 × 10⁷ cells) at 5 days after irradiation. Virus stocks were prepared from 24-h harvests of culture supernatants at day 9, and HIV-1 p24 production (determined by enzyme-linked immunosorbent assay (ELISA)) is shown (PBL) for virus stocks used in experiments *a, b* and *c*. In study 2, virus stocks were prepared from 24-h harvests at days 5–8; representative HIV-1 p24 production in PBL at day-7 is shown. Recovery of the recombinant strains in study 1 was confirmed by Southern hybridization analysis of infected PBL. Blood mononuclear phagocytes were purified by adherence to culture dishes in Iscove's medium, 5% human AB serum (Pel-Freez) and 15% fetal calf serum for 3 h. Adherent cells were washed twice and propagated in Iscove's medium, 5% human AB serum and 15% FCS for an additional 24 h before propagation in Iscove's medium and 20% FCS for 6 days. Cells were treated with 0.5 mM EDTA in PBS, replated in 24-well plates (Falcon) at 5 × 10⁵ cells per well, and exposed to HIV-1 p24 (10 ng per well in study 1, and 50 ng per well in study 2) for 2 h in the presence of 10 µg ml⁻¹ polybrene, then washed twice. Medium was changed every 3–4 days, and culture supernatants assayed for p24 antigen (ELISA, Abbott) at days 14 and 21.

NT, not tested.

isolated after infection of PBL. The complete nucleotide sequence of HIV-1_{JR-FL} was determined (Y.K. and I.S.Y.C., manuscript in preparation) to allow precise construction of a series of molecular recombinants between HIV-1_{JR-FL} and HIV-1_{NL4-3} using conserved restriction enzyme sites (Fig. 1*a*).

Recombinant virus strains containing sequences from regions of the 3' end of HIV-1_{JR-FL} that were made progressively smaller by use of shared restriction enzyme sites (*Sst*I at nucleotide 5,999; *Dra*III at nucleotide 6,591; and *Stu*I at nucleotide 6,822) were able to infect mononuclear phagocytes productively. But replication was inefficient when mononuclear phagocytes were infected with a recombinant virus, HIV-1_{NF-Xho} (containing sequences from HIV-1_{JR-FL} downstream of *Xho*I (nucleotide 8,887), including *nef* and U3/R elements of the HIV-1_{JR-FL} long terminal repeat). Thus, *cis*-acting regulatory sequences in the long terminal repeat do not seem to be important for macrophage tropism. To confirm that the region between nucleotides 6,822 and 8,887 is necessary for replication in mononuclear

phagocytes, we substituted an internal fragment into pNL4-3. This recombinant HIV-1_{NFN-SX}, which contained sequences from HIV-1_{JR-FL} between the *Stu*I site and *Xho*I site (2.1 kilobase (kb)), replicated as efficiently in blood mononuclear phagocytes as the parental strain, HIV-1_{JR-FL} (Table 1, experiments *a* to *f*). Kinetics of virus production from experiment *a* are represented in Fig. 1*b*: the results show that mononuclear phagocyte tropism is genetically determined, and that the sequences responsible lie within a 2.1-kb region at the 3' end of the virus. This region includes part of the envelope gene, the second coding exon of *tat* and *rev*, and a small portion of the *nef* gene.

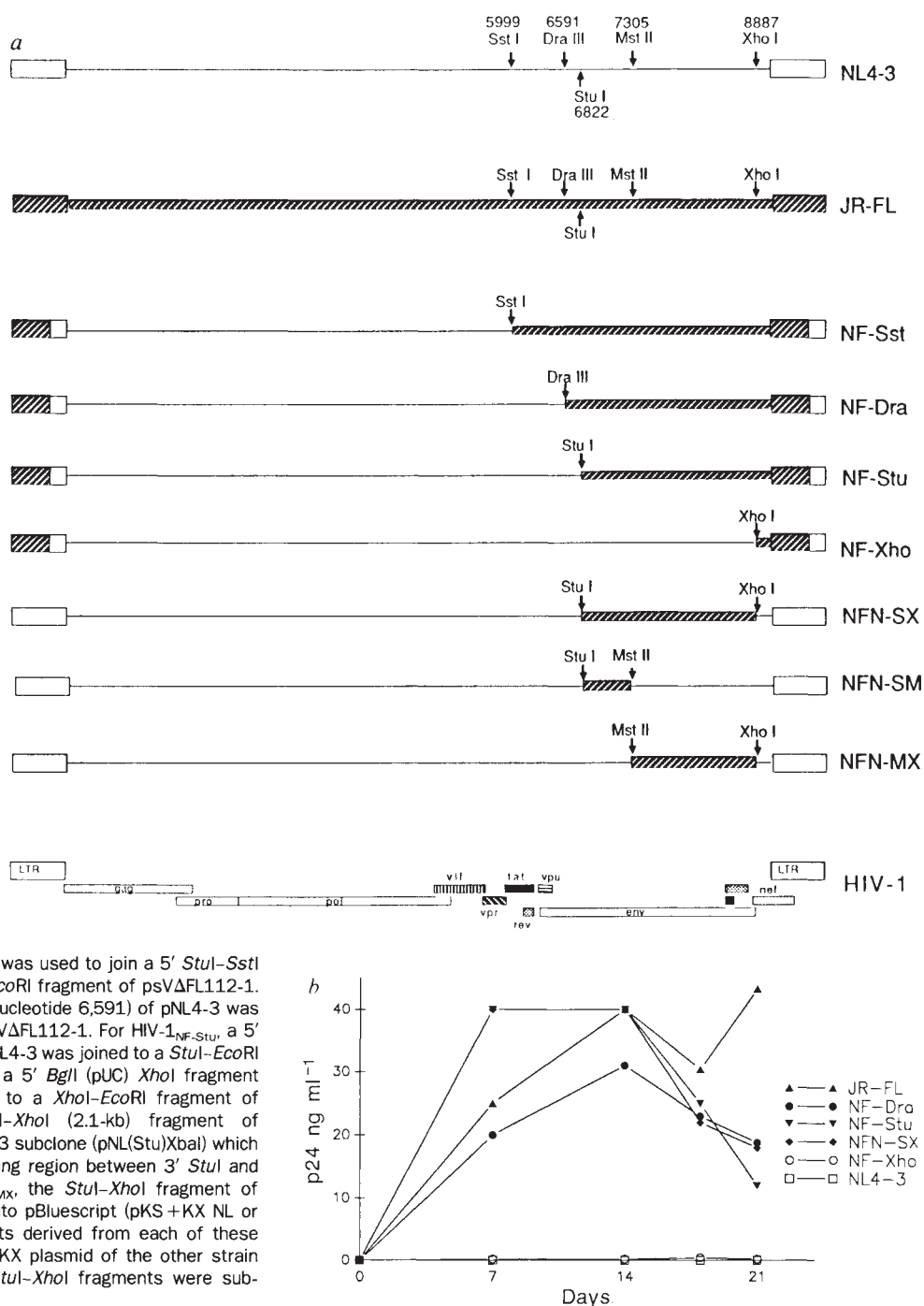
More recombinant strains were generated by substitution of smaller fragments of HIV-1_{JR-FL} from within the 2.1-kb *Stu*I-*Xho*I region using the shared restriction enzyme site, *Mst*II (nucleotide 7,305). The recombinant HIV-1_{NFN-SM}, which contains 157 amino acids from the HIV-1_{JR-FL} gp120 region (residues 202 to 358) efficiently infects mononuclear phagocytes (Table 1). This recombinant does not include the previously defined

FIG. 1 Construction of recombinant virus isolates and replication in blood mononuclear phagocytes. Restriction enzyme sites *Sst*I (nucleotide 5,999), *Dra*III (nucleotide 6,591), *Stu*I (nucleotide 6,822), and *Xho*I (nucleotide 8,887) were identified by nucleotide sequencing of cloned HIV-1_{JR-FL} and HIV-1_{NL4-3} (ref. 21) DNA, and used to prepare hybrid recombinant genomes (a). The proviral DNA predicted to result from transfection of ligated fragments is shown. Cross-hatched lines and boxes depict sequences derived from HIV-1_{JR-FL}, and simple lines and open boxes depict sequences derived from HIV-1_{NL4-3}. The positions of the HIV-1 genes are shown at the bottom of the panel. LTR, Long terminal repeat kinetics of replication (b) of HIV-1 parental and recombinant strains of a representative experiment (see Table 1, a) are shown for HIV-1_{JR-FL} (▲), HIV-1_{NF-Dra} (●), HIV-1_{NF-Stu} (▼), HIV-1_{NF-SX} (◆), HIV-1_{NF-Xho} (○) and HIV-1_{NL4-3} (□). Media were changed every 3–4 days and assayed for virus at days 7, 14, 18 and 21.

METHODS. The unintegrated form of HIV-1_{JR-FL} DNA was cloned into the *Sst*I site of vector λ ongC. The *Sst*I-*Sst*I fragment (nucleotides 5,999-9,574) was subcloned into the *Sst*I site of a complete single long terminal repeat of HIV-1_{JR-CSF}, and flanked by cellular DNA sequences joined at *Eco*RI to the vector pSV2neo (pSV Δ FL112-1). Thus, the *R* region of the resulting LTR is also recombinant for HIV-1_{JR-FL} and HIV-1_{JR-CSF}. The hybrid proviral DNAs were constructed as follows. For HIV-1_{NF-Sst}, a *Stu*I site in the 5' flanking cellular sequences of the infectious clone, pNL4-3, was used to join a 5' *Stu*I-*Sst*I fragment (nucleotide 5,999) to a 3' *Sst*I-*Eco*RI fragment of pSV Δ FL112-1. For HIV-1_{NF-Dra}, a 5' *Stu*I-*Dra*III fragment (nucleotide 6,591) of pNL4-3 was joined to a 3' *Dra*III-*Eco*RI fragment of pSV Δ FL112-1. For HIV-1_{NF-Stu}, a 5' *Stu*I-*Stu*I fragment (nucleotide 6,822) of pNL4-3 was joined to a *Stu*I-*Eco*RI fragment of pSV Δ FL112. For HIV-1_{NF-Xho}, a 5' *Bgl*II (pUC) *Xho*I fragment (nucleotide 8,887) of pNL4-3 was joined to a *Xho*I-*Eco*RI fragment of pSV Δ FL112-1. For HIV-1_{NFN-SX}, a *Stu*I-*Xho*I (2.1-kb) fragment of pSV Δ FL112-1 was substituted into a pNL4-3 subclone (pNL(Stu)*Xba*I) which had sequences removed from the 5' flanking region between 3' *Stu*I and *Xho*I sites. For HIV-1_{NFN-SM} and HIV-1_{NFN-MX}, the *Stu*I-*Xho*I fragment of pSV Δ FL112-1 or pNL4-3 was subcloned into pBluescript (pKS+KX NL or pKS+KX FL). *Stu*I-*Mst*II (0.5-kb) fragments derived from each of these plasmids were substituted into the pKS+KX plasmid of the other strain (pKS+NF *Mst* or pKS+FN *Mst*). Hybrid *Stu*I-*Xho*I fragments were substituted into pNL(Stu)*Xba*I.

CD4 binding domain¹⁹, namely amino acids 404 to 447. Replication in mononuclear phagocytes was not efficient when we used the recombinant HIV-1_{NFN-MX}, which contains the CD4-binding domain and all of glycoprotein gp41 encoded by HIV-1_{JR-FL}. Therefore at least one genetic determinant that is sufficient to confer mononuclear phagocyte tropism must reside in gp120, upstream and distinct from the CD4-binding domain. It is notable that HIV-1_{NFN-SM} seems to replicate rather less efficiently than HIV-1_{NFN-SX}, suggesting that other regions may also play a part in determining tropism.

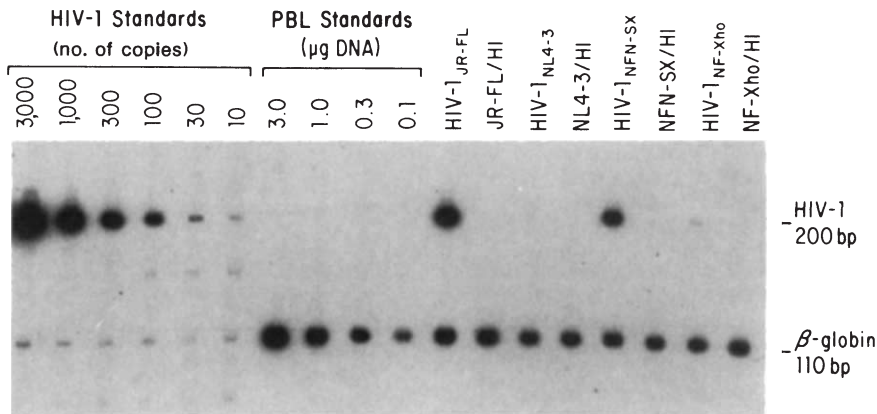
We further investigated the mechanism behind these tropism differences by assaying HIV-1-specific DNA formation early after infection by adapting a modification of the polymerase chain reaction (PCR) procedure^{22,23}, which is both sensitive and quantitative²⁴⁻²⁸. Peripheral blood mononuclear phagocytes were infected with the same amount of infectious virus (as determined by limiting-dilution assays in PBL) using parental



strains HIV-1_{JR-FL} and HIV-1_{NL4-3}, and two recombinants, HIV-1_{NFN-SX} and HIV-1_{NF-Xho}. Comparable amounts of HIV-1-specific DNA were detected after infection with HIV-1_{JR-FL} and HIV-1_{NFN-SX} (800 copies per μg of cellular DNA and 1,000 copies per μg of cellular DNA, respectively; note the copy numbers are normalized by the β -globin signal) (Fig. 2). Much less DNA (10 copies per μg) relative to the heat-inactivated control was detected after infection with HIV-1_{NL4-3} and HIV-1_{NF-Xho}. In agreement with our *env* recombinant results, this demonstrates that the strains that are poorly tropic for macrophages enter mononuclear phagocytes less efficiently than the strains that can replicate in mononuclear phagocytes. PCR probes designed to detect any initial reverse transcripts²⁷ give similar results. These observations parallel our studies of restricted T-cell line tropism²⁸.

Others have claimed that sequences that include the *env* gene can affect viral tropism²⁹⁻³¹. One report²⁹ concludes that muta-

FIG. 2 Quantitative detection of HIV-1 DNA and cellular DNA by PCR. Normal mononuclear phagocytes (5×10^6) purified by adherence were infected on day 6 at a multiplicity of infection 0.05, as described in Table 1. Heat-inactivated (HI) virus supernatants were used to infect cells in parallel. Cells were collected for isolation of total nucleic acids, and $\sim 1.0 \mu\text{g}$ DNA from each sample was subjected to 25 cycles of PCR. Serial threefold dilutions of cloned HIV-1 DNA from 10^4 to 10 copies were diluted in carrier DNA (uninfected PBL DNA at a concentration of $5 \mu\text{g ml}^{-1}$), and serial threefold dilutions of HIV-1-negative total cellular DNA from $3.0 \mu\text{g}$ to 100 ng were subjected in parallel to 25 cycles of PCR. METHODS. Contaminating viral DNA in virus supernatants was removed by DNase I treatment²⁷. Heat-inactivated virus supernatants were treated at 60°C for 1 h. Infected cells were washed once in phosphate-buffered saline, and lysed in 0.5% (w/v) SDS. Proteins were digested with proteinase K ($200 \mu\text{g ml}^{-1}$) at 50°C for 3 h, followed by extraction with phenol/chloroform and precipitation with ethanol. Total nucleic acids (DNA + RNA) from $\sim 10^5$ cells were subjected to 25 PCR cycles of 1 min at 91°C for denaturation and 2 min at 65°C for annealing and extension. The HIV-1 LTR/*gag*-specific oligonucleotide pair, M661 and M667



(200-bp product)²⁷, and the globin-specific oligonucleotide pair (110-bp product)^{22,23} were included together for each sample. Amplified products were resolved by electrophoresis on an 8% acrylamide gel and analysed directly by autoradiography. Quantitation was achieved by measuring c.p.m. from excised bands corresponding to the specific amplified product, and comparing this with a standard curve generated from HIV-1 and PBL controls.

tions in the CD4-binding region, corresponding to amino-acid positions 411 and 418 of HIV-1_{JR-FL}, alters tropism of U937 and Sup T1 cells. Those amino acids are conserved between HIV-1_{JR-FL} and HIV-1_{NL4-3}, and in most strains of HIV-1. In addition, as HIV-1_{JR-FL} does not productively infect U937 cells whereas HIV-1_{NL4-3} does, tropism in primary mononuclear phagocytes probably more accurately represents tropism *in vivo*. Also, the earlier studies^{30,31} had ill-defined endpoints of recombination which included large regions of *gp120*, *gp41*, *vpu*, *rev*, *tat* and *nef*. We exclude genes other than *env*, and also the known functional domains of *env*, the CD4-binding domain in *gp120* (ref. 19), and the fusion domain in *gp41* (ref. 32), which could be critical for entry.

The deduced amino-acid sequence of the entire *env* gene from HIV-1_{JR-FL} (Fig. 3) differs by 12% from that of HIV-1_{NL4-3}. The amino-acid sequence of *env* between the *StuI* and *MstII* sites (amino acids 202 to 358) differs by 13.9%. Several domains defined by neutralization or mutagenesis have been characterized in this region of *gp120*. The V3 neutralization domain, amino acids 293 to 327, is thought to form a loop structure by disulphide linkage of two cysteine residues, and is the principal immunogenic domain for type-specific neutralization (reviewed in ref. 33). Deletion or mutation of the loop region reduces infectivity without affecting binding to CD4. Antisera against synthetic peptides homologous to the second conserved region of *gp120* (residues 244 to 264 of HIV-1_{JR-FL})³⁴ or a point mutation (amino acid 259 of HIV-1_{JR-FL})³⁵ did not block CD4 binding, but abrogated infectivity of some HIV-1 strains in T-cells. Deletion and insertion mutations in this region can also disrupt CD4 binding^{32,36} or fusion³², suggesting that interactions between various *env* domains are important in virus entry.

It is notable that, regardless of their ability to replicate in mononuclear phagocytes, all parental and recombinant strains replicate equally efficiently in activated PBL (Table 1). As soluble CD4 and Leu 3a monoclonal antibody block both PBL

and mononuclear phagocyte infection by HIV-1_{JR-FL} (D. Ho, personal communication), CD4 is clearly a critical determinant of HIV-1_{JR-FL} infection of cells. On the basis of our results, we propose that target cell interactions with the region of *gp120*



FIG. 3 Amino-acid sequence deduced from the entire *env* gene of HIV-1_{NL4-3} and HIV-1_{JR-FL}. The numbering system is based on the sequence of *gp120/gp41* from molecular clones of HIV-1_{NL4-3} or HIV-1_{JR-FL}. Functional domains of *env* are indicated. Amino-acid sequences within the *StuI*-*XhoI* region are shaded. Comparison of the sequences was performed by the 'best-fit' method³⁸. Dashes indicate that the corresponding HIV-1_{JR-FL} amino acid is identical to that in HIV-1_{NL4-3}. Amino-acid substitutions are indicated, and deletions are indicated by a single dot.

spanning amino acids 202 to 358, in addition to or in conjunction with the CD4-binding domain, are necessary for efficient entry of virus into mononuclear phagocytes.

HIV-1 strains that can replicate efficiently in mononuclear phagocytes have been recovered primarily from neural and pulmonary tissues of infected individuals^{17,18}. In particular, AIDS dementia is highly correlated with infection of macrophages and/or microglia in the brain, and HIV-1_{JR-FL} was isolated from brain tissue of a patient who died with severe AIDS dementia. Our studies suggest a possible link between genetic properties of HIV-1 and an important disease manifestation of AIDS, AIDS dementia. □

Received 29 June; accepted 29 August 1990.

- Koenig, S. *et al. Science* **233**, 1089–1093 (1986).
- Gabuzda, D. H. *et al. Ann. Neurol.* **20**, 289–295 (1986).
- Wiley, C. A., Schrier, R. D., Nelson, J. A., Lampert, P. W. & Oldstone, M. B. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7089–7093 (1986).
- Gartner, S., Markovits, P., Markovits, D. M., Betts, R. F. & Popovic, M. *J. Am. med. Assn* **256**, 2365–2371 (1986).
- Ellbott, D. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 3337–3341 (1989).
- Chayt, K. J. *et al. J. Am. med. Assn* **256**, 2356–2359 (1986).
- Tschachler, E. *et al. J. Invest. Dermatol.* **88**, 233–237 (1987).
- Rappersberger, K. *et al. Intervirol* **29**, 185–194 (1988).
- Levy, R. M., Bredeisen, D. E. & Rosenblum, M. L. *J. Neurosurg.* **62**, 475–495 (1985).
- Navia, B. A., Jordan, B. D. & Price, R. W. *Ann. Neurol.* **19**, 517–524 (1986).

- Petito, C. K. *et al. New. Engl. J. Med.* **312**, 874–879 (1985).
- Lipkin, W. I., Parry, G., Kiprov, D. & Abrams, D. *Neurology* **35**, 1479–1483 (1985).
- Cornblath, D. R., McArthur, J. C., Kennedy, P. G., Witte, A. S. & Griffin, J. W. *Ann. Intern. Med.* **21**, 32–40 (1987).
- Suffredini, A. F. *et al. Ann. Intern. Med.* **107**, 7–13 (1987).
- Johnson, T. M., Duvic, M., Rapini, R. P. & Rios, A. *New Engl. J. Med.* **313**, 1415 (1985).
- Eisenstat, B. A. & Wormser, G. P. *New Engl. J. Med.* **311**, 189 (1984).
- Gartner, S. *et al. Science* **233**, 215–219 (1986).
- Koyanagi, Y. *et al. Science* **236**, 819–822 (1987).
- Lasky, L. A. *et al. Cell* **50**, 975–985 (1987).
- Levy, J. A. *et al. Science* **225**, 840–842 (1984).
- Adachi, A. *et al. J. Virol.* **59**, 284–291 (1986).
- Saiki, R. K. *et al. Science* **230**, 1350–1354 (1985).
- Saiki, R. K. *et al. Science* **239**, 487–491 (1988).
- Lee, H. *et al. Science* **244**, 471–475 (1989).
- Pang, S. *et al. Nature* **343**, 85–90 (1990).
- Arrigo, S. J., Weitsman, S., Rosenblatt, J. D. & Chen, I. S. Y. *J. Virol.* **63**, 4875–4881 (1989).
- Zack, J. A. *et al. Cell* **61**, 213–222 (1990).
- Cann, A. J. *et al. J. Virol.* **64**, 4733–4742 (1990).
- Cordonnier, A., Montagnier, L. & Emerman, M. *Nature* **340**, 571–574 (1989).
- Fisher, A. G. *et al. Nature* **334**, 444–447 (1988).
- Cheng-Mayer, C. *et al. J. Virol.* **64**, 4390–4398 (1990).
- Kowalski, M. *et al. Science* **237**, 1351–1355 (1987).
- Bolognesi, D. P. *Molec. biol. Med.* **7**, 1–15 (1990).
- Ho, D. D., Kaplan, J. C., Rackauskas, I. E. & Gurney, M. E. *Science* **239**, 1021–1023 (1988).
- Willey, R. L. *et al. J. Virol.* **62**, 139–147 (1988).
- Cordonnier, A., Montagnier, L. & Emerman, M. *J. Virol.* **63**, 4464–4468 (1989).
- Cann, A. J., Koyanagi, Y. & Chen, I. S. Y. *Oncogene* **3**, 123–128 (1988).
- Devereux, J., Haebesli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1986).

ACKNOWLEDGEMENTS. Supported by NIH and Universitywide Task Force on AIDS. I.S.Y.C. is supported in part by a Leukemia Society of America Scholar Award. We thank M. Martin for the pNL4-3 clone; S. Arrigo, P. Green, D. Ho, V. Miller and Shen Pang for critical reading of the manuscript; S. Lee for technical assistance; and W. Aft for preparation of the manuscript.

Analysis of a human DNA excision repair gene involved in group A xeroderma pigmentosum and containing a zinc-finger domain

Kiyoji Tanaka*, Naoyuki Miura*, Ichiro Satokata*, Iwai Miyamoto*, Michihiro C. Yoshida†, Yoshiaki Satoh‡§, Seiji Kondo§, Akira Yasui||, Hiroto Okayama¶ & Yoshio Okada*

* Institute for Molecular and Cellular Biology,

¶ Research Institute for Microbial Diseases, Osaka University, Osaka 565, Japan

† Chromosome Research Unit, Faculty of Science, Hokkaido University, Sapporo 060, Japan

‡ Department of Dermatology, Daini-Hospital, Tokyo Women's Medical College, Tokyo 116, Japan

§ Department of Dermatology, Tokyo Medical and Dental University, Tokyo 113, Japan

|| Research Institute for Tuberculosis and Cancer, Tohoku University, Sendai 980, Japan

XERODERMA pigmentosum (XP) is an autosomal recessive disease, characterized by a high incidence of sunlight-induced skin cancer. Cells from people with this condition are hypersensitive to ultraviolet because of a defect in DNA repair. There are nine genetic complementation groups of XP, groups A–H and a variant. We have cloned the mouse DNA repair gene that complements the defect of group A, the *XPAC* gene¹. Here we report molecular cloning of human and mouse *XPAC* complementary DNAs. Expression of *XPAC* cDNA confers ultraviolet-resistance on several group A cell lines, but not on lines of other XP groups. Almost all group A lines tested showed abnormality or absence of *XPAC* messenger RNAs. These results indicate that a defective *XPAC* gene causes group A XP. The human and mouse *XPAC* genes are located on chromosome 9q34.1 and chromosome 4C2, respectively. Human *XPAC* cDNA encodes a protein of 273 amino acids with a zinc-finger motif.

We cloned the *XPAC* gene, and detected 1.0–1.1 kilobase (kb) *XPAC* mRNA in mouse cells, and 1.3–1.4 kb and 1.0–1.1 kb *XPAC* mRNAs in normal human cells¹. To obtain human and

mouse cDNAs corresponding to these mRNAs, we screened pcD2 human and mouse expression cDNA libraries² using one of the exon-containing mouse genomic fragments as a probe¹. The positive cDNA clones obtained were then transfected into group A XP cells (XP2OSSV) by the calcium phosphate coprecipitation method³. As the pcD2 vector contains the *Tn5* phosphotransferase (*neo*) gene, XP transfectants were initially selected by resistance to G418. They were then irradiated with ultraviolet (4 Jm⁻²), with a 4-day interval between irradiations.

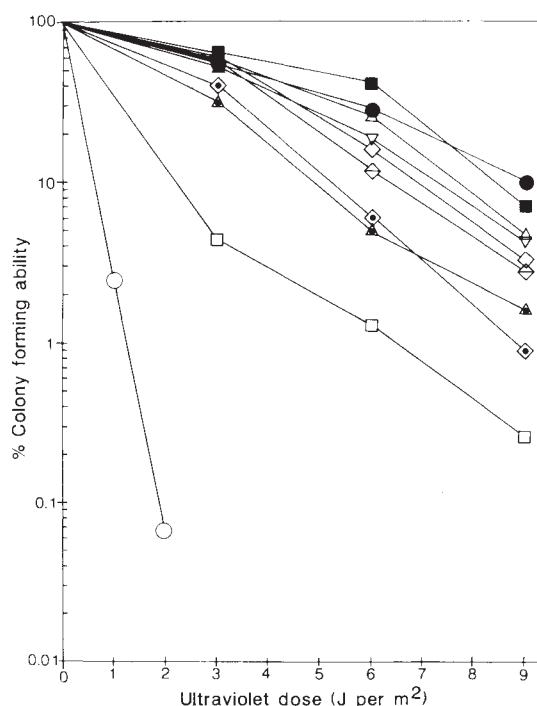
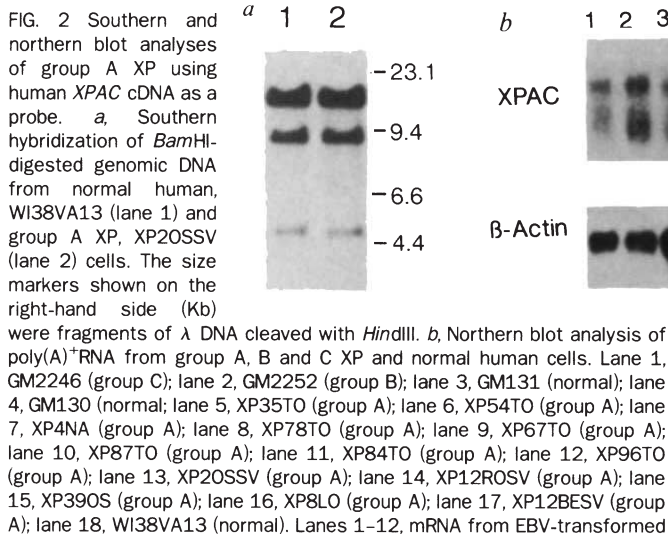


FIG. 1 Ultraviolet-sensitivity, measured by colony-forming ability of GM637 (normal human, ●), WI38VA13 (normal human, ○), XP2OSSV (group A XP, ▲), ultraviolet-resistant pcD2h64-transfected XP2OSSV (◆), ultraviolet-resistant pcD2h19-transfected XP2OSSV (△, ▽), XP12BESV (group A XP, □), ultraviolet-resistant pcD2h64-transfected XP12BESV (▲), ultraviolet-resistant pcD2h19-transfected XP12BESV (◇, ◇) cells. All the transfectants were the individual clones. Points are average values for duplicate dishes. Ultraviolet-survival curves were examined as described¹.



After transfection of pcD2m13 or 95 (mouse), pcD2h19, 29, 64 or 67 (human) cDNA clones into XP2OSSV cells, about one-third of the G418-resistant XP transfectants showed ultraviolet-resistance. One of these human *XPAC* cDNA clones, pcD2h19, was then transfected into several group A XP cell lines (XP2OSSV, XP12ROSV and XP12BESV) and lines of other XP complementation groups (XP4PASV, XP6BESV, XP2YOSV and XP3BRSV; group C, D, F and G, respectively). About 20–30% of G418-resistant transfectants of every group A XP cell line showed ultraviolet-resistance, but no ultraviolet-resistant transfectants were obtained from XP cells of other groups, confirming that the recovery of DNA repair ability by the *XPAC* gene is specific for group A (data not shown). Figure 1 shows the ultraviolet-survival curves of the ultraviolet-resistant group A XP transfectants. The group A cells transfected with pcD2h19 seemed to be more ultraviolet-resistant than those transfected with pcD2h64, but more ultraviolet-sensitive than normal human cells.

The *XPAC* genome and mRNA in people with group A XP were then subjected to Southern and northern blot analyses with human *XPAC* cDNA (pcD2h64 insert) as a probe. Samples of high-molecular-weight genomic DNA from 23 different group A patients (20 of whom were Japanese) were digested with either *Bam*HI, *Pst*I, *Eco*RI, *Hind*III or *Taq*I, separated on 0.7% agarose gels and blot-hybridized with *XPAC* cDNA. The results in Fig. 2*a* show that there was no gross genomic rearrangement of the *XPAC* gene in any group A patient tested. Northern blot analysis of poly(A)⁺RNA revealed *XPAC* mRNA of about 1.3–1.4 kb and 1.0–1.1 kb in cells of groups B–H and a variant XP and in normal human cells, but abnormal *XPAC* mRNAs in almost all group A patients (Fig. 2*b*).

Almost all the Japanese samples (except XP39OS, Fig. 2*b* lane 15) showed identical abnormalities in *XPAC* mRNAs (reduced amounts of about 1.2–1.3 kb and 0.9–1.0 kb *XPAC* mRNAs) in northern blot analysis, indicating a common origin of mutation of the *XPAC* gene in almost all our Japanese patients. This mutation is a G→C transversion at the 3' splice acceptor site of intron 3 (I.S., K.T., N.M., I.M., Y.S., S.K., Y.O., manuscript submitted). In XP12ROSV and XP12BESV (both non-Japanese) cells, no significant amount of *XPAC* mRNA was detected by northern blot analysis (Fig. 2*b*, lanes 14 and 17). In XP39OS (Japanese) and XP8LO (non-Japanese) group A XP patients, who showed mild skin symptoms with few, if any, skin tumours or neurological abnormalities^{4,5}, normal *XPAC* mRNAs were detected (Fig. 2*b* lanes 15 and 16). These patients have a missense or nonsense mutation in the *XPAC* gene (I.S. *et al.* unpublished data). These findings indicate that the *XPAC* gene is a cause of group A XP and that different mutations in the *XPAC* gene can result in group A XP.

lymphoblastoid cells. Lanes 13–18, mRNA from SV40-transformed fibroblast cells. Sizes of main species (Kb) on right-hand side.

METHODS. Some XP cells used in this study have been described elsewhere^{1,19}. XP8LO cells were provided by J. H. J. Hoeijmakers. XP39OS cells were obtained from the JCRB Cell Bank. Other XP cells were newly established for this study. DNA was extracted, digested with restriction enzymes and blot-hybridized with human *XPAC* cDNA as described¹. Poly(A)⁺RNA was purified and blot-hybridized with human *XPAC* cDNA as described¹.

The human and mouse *XPAC* genes were assigned to positions 9q34.1 and 4C2, respectively by chromosomal *in situ* hybridization (Fig. 3*a, b*). This is consistent with the results obtained previously by chromosomal transfer^{6–8}, and provides additional evidence that the *XPAC* gene is a cause of group A XP. Comparative mapping studies showed that human chromosome 9 has homologous loci to mouse chromosomes 2, 4 and 19; human 9q34 corresponds to mouse 4C2, on which a homologous gene for Δ -aminolevulinate dehydrogenase is located⁹. Our data presented here demonstrate another homologous gene on human 9q34 and mouse 4C2, but it is not known whether the *XPAC* gene is syntenic to the Δ -aminolevulinate dehydrogenase gene.

The nucleotide sequences of inserts of human and mouse *XPAC* cDNA clones were determined by the dideoxynucleotide

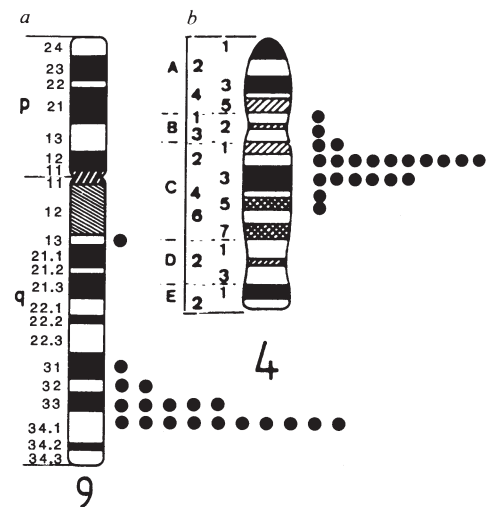


FIG. 3 Chromosomal localizations of human and mouse *XPAC* genes by *in situ* hybridization. *a*, Idiogram showing the distribution of silver grains along chromosome 9 with a peak at q34.1. *b*, Idiogram showing the silver grain distribution on mouse chromosome 4 with a peak at C2.

METHODS. Human chromosome preparations were obtained from phytohaemagglutinin-stimulated peripheral lymphocytes. Mouse chromosome metaphases were prepared from primary embryonic cultured fibroblasts derived from the DDD strain. For *in situ* hybridization experiments, human and mouse *XPAC* cDNAs were labelled by either the nick-translation or random primer procedure in the presence of ³H-labelled nucleotide triphosphates and hybridized to chromosome preparations of human or mouse metaphases, which were subsequently processed for autoradiography as described²⁰. The mouse chromosomes were identified by the standard numbering system^{21,22}.

chain-termination method¹⁰. The human cDNA clone pcD2h19 has a single large open reading frame predicted to encode 273 amino-acid residues, starting at the initiation codon 27 nucleotides downstream from the 5' end and ending with a TGA termination codon at position 820 (Fig. 3a, b). The sequences around the predicted initiator methionine codon agree with the consensus sequences described by Kozak¹¹. Our preliminary results indicate that the transcription start sites are 156 nucleotides upstream from the initiation methionine codon, and that there are no other in-frame ATG codons between the transcription start sites and the initiation codon (I.S. *et al.*, unpublished data). On the other hand, sequence analysis revealed that the clones pcD2h29, 64 and 67 have the same XPAC cDNA insert, but they lack the 5' part including the translational start and 58 N-terminal amino acids (Fig. 4a, b). These results indicate that

a truncated XPAC protein that may be translated from an ATG present at position 176 retains sufficient function to restore partially the DNA repair defect in group A XP cells. The relative molecular mass (M_r) of the human XPAC protein calculated from the XPAC cDNA insert of pcD2h19 is about 31,000 (31K). The M_r of the XP A correcting protein partially purified from HeLa cells or calf thymus was about 40–45K on SDS-PAGE gels (J. H. J. Hoeijmakers and M. Yamaizumi, personal communication). Our preliminary data indicate that two XPAC proteins of about 40 and 42K are detectable in normal human cells on immunoprecipitation using polyclonal antibody against recombinant human XPAC protein. Two XPAC proteins of the same sizes were also detected in ultraviolet-resistant group A XP cells transfected with pcD2h19, but absent in group A XP cells; also, the production of XPAC protein was less in these

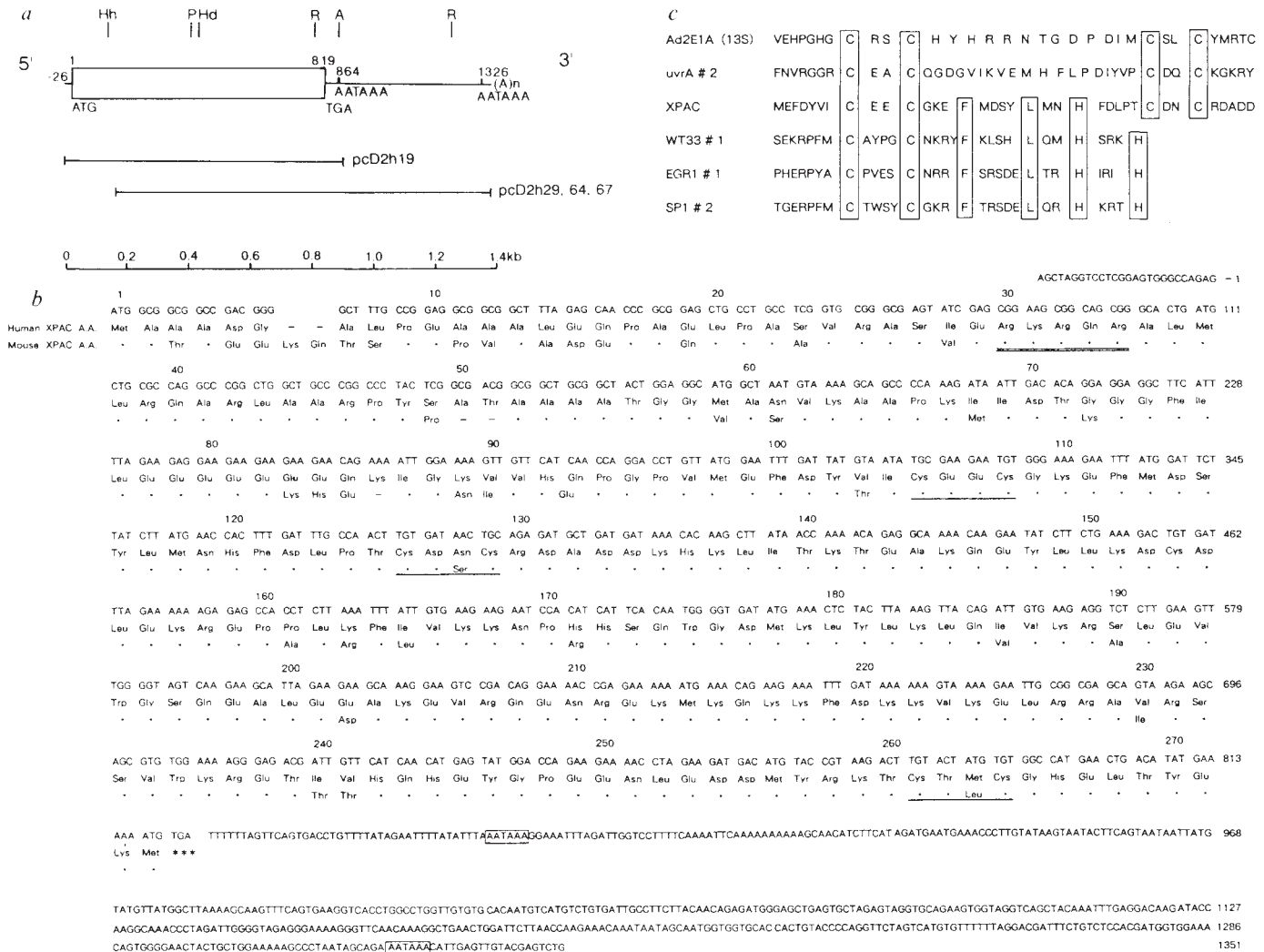


FIG. 4 Restriction map, nucleotide sequence and deduced amino acid sequence of XPAC cDNA. **a**, Structure and restriction map of human XPAC cDNA. The untranslated sequences and coding sequences are represented by lines and a box, respectively. Horizontal bars represent the inserts of four human cDNA clones, pcD2h19, 29, 64 and 67. Hh, *HhaI* site; P, *PstI* site; Hd, *HindIII* site; R, *RsaI* site; A, *AhaII* site. **b**, Nucleotide sequences of human XPAC cDNA with deduced amino acid sequence of the open reading frame. The amino acid sequence in the second row is that of the mouse XPAC protein. Dots in the mouse XPAC amino acid sequence indicate amino acids conserved in the human XPAC protein. Hyphens indicate gaps introduced to optimize matching between human and mouse XPAC amino acid sequences. The zinc-finger motif, CX₂C is underlined. The putative nuclear localization signal is indicated by a double bar. Poly(A)-addition signal sequences are boxed. Asterisks mark the termination codon. **c**, Amino-acid sequences of the putative human XPAC-associated zinc-finger domain and comparison with the amino-acid sequences of other C₂H₂ and C₄ class zinc-

fingers; that is, encoded by adenovirus type 2 E1A gene²³, *uvrA* gene¹⁰, *WT33* (Wilm's tumour gene)²⁴, *EGR1* (early growth response gene)²⁵ and the SP1 protein²⁶. Boxes indicate the CX₂C or HX₂H zinc-finger motif and amino acids conserved in XPAC and C₂H₂ class zinc-fingers. **METHODS.** Human and mouse expression pcD2 cDNA libraries were prepared from poly(A)⁺ RNA of human primary fibroblasts and methylcholanthrene-transformed C3H10T1/2 cells², transfected into *E. coli* BH101, and screened by hybridization with a ³²P-labelled 0.7 kb subfragment of the B₇-B₈ fragment of EMBL3/24B¹. About 20 positive clones were obtained from each library with a frequency of one per 2 × 10⁵ clones. Transfections of some of these positive clones into group A XP cells conferred ultraviolet-resistance on XP cells (pcD2m13 and 95, pcD2h19, 29, 64 and 67). The cDNA inserts of these clones were subcloned into M13mp18 and M13mp19, and sequenced by the dideoxy-termination method⁶ with Sequenase version 2.0 (US Biochemicals).

transfectants than in normal human cells (N.M. *et al.*, unpublished data). These results suggest that the XPAC protein migrates slowly in gels because there are several proline residues in the N-terminal half of XPAC protein and because of the internal acidic domain. It seems that the pcD2h19 clone encodes the full-length human XPAC protein and the incomplete recovery of ultraviolet-resistance of the group A XP cells transfected with pcD2h19 might be due to a lower amount of XPAC protein. Two potential polyadenylation signals (AATAAA) occur at nucleotide positions 864 and 1,326 (Fig. 4a, b). This finding suggests that the two XPAC mRNAs in normal human cells are due to alternative polyadenylations and that the pcD2h19 clone might represent 1.0–1.4 kb XPAC mRNA, while pcD2h29, 64 and 67 might represent 1.3–1.4 kb XPAC mRNA (Fig. 4a). In mouse cells, only 1.0–1.1 kb XPAC mRNA was detected, and both pcD2m13 and 95 retain only one polyadenylation signal, 23 nucleotides downstream from the TGA termination codon (data not shown). The 5' region of XPAC cDNA is GC-rich, whereas the 3' untranslated region is AT-rich. The XPAC protein is rich in glutamic acid (12.4%), lysine (10.6%) and alanine (9.9%) and has low hydrophobicity (data not shown). No proteins with extensive sequence similarity with the XPAC protein were found in a search of the National Biomedical Research Foundation database. A mouse XPAC protein was 95% similar to a human XPAC protein (Fig. 4b). A computer search of the secondary structure using the program of Chou and Fasman¹² revealed several α -helix-turn-helix structures in the XPAC protein (for example, centred at amino-acid residues 95, 152, 196 and 231). In addition, the XPAC protein contains a zinc-finger motif, CX₂CX₃FX₄LX₂HX₃CX₂C (single letter amino-acid code), at amino acid positions 105–129. This finding strongly suggests that the XPAC protein is a DNA-binding protein. In general, there are two classes of zinc-finger proteins: the Cx family (for example, adenovirus E1A and the steroid hormone receptor superfamily) and the C₂H₂ family (for example, SP1 and TFIIIA)¹³. In the C₂H₂ family, the finger motif is tandemly repeated (2–13 times) with a 7 or 8 amino-acid linker separating the units, whereas the Cx proteins have nonrepetitive fingers. There is another CX₂C sequence at amino-acid positions 261–264 in the XPAC protein, but no repeat of the zinc-finger unit. In this sense, the XPAC protein belongs to the Cx(C₄) family. The amino acid sequences conserved in a loop structure between pairs of cysteines and histidines in the C₂H₂ family are also well conserved in the XPAC zinc-finger (Fig. 4c). Thus, the XPAC zinc-finger has motifs of both the C₄ and C₂H₂ protein families.

The uvrA protein and poly(ADP-ribose)polymerase, which are involved in DNA repair in *Escherichia coli* and mammalian cells, respectively, both contain a zinc-finger motif and bind to DNA^{14–16}. The XPAC protein is a third DNA repair protein with a zinc-finger motif. The uvrA protein, like the XPAC protein, is involved in the incision step of excision repair. The zinc-finger motif in the uvrA protein also belongs to the C₄ family, but the uvrA protein contains two units of C₄ because of internal gene duplication (two fingers). Like the XPAC protein, the uvrA protein also contains other unpaired CX₂C sequences. These results suggest that the function of the XPAC protein in the incision step of excision repair might be similar to that of the uvrA protein. The uvrA protein has ATPase activity and ATP stimulates its binding to ultraviolet-irradiated DNA^{17,18}. But the XPAC protein does not seem to have a nucleotide-binding domain or helicase motif, indicating that its function is not the same. The binding sites for the Cx class protein are relatively short and show dyad symmetry, indicating dimer formation for DNA-binding¹³. It is possible that the XPAC protein forms a homodimer for DNA binding. □

3. Corsaro, C. M. & Pearson, M. L. *Somatic Cell Genet.* **7**, 603–616 (1981).
4. Sato, K. *et al.* *Br. J. Derm.* **116**, 101–108 (1987).
5. de Weerd-Kastelein, E. A. *et al.* *Mutat. Res.* **37**, 307–312 (1976).
6. Lin, P. & Ruddle, F. H. *Nature* **289**, 191–194 (1981).
7. Kaur, G. P. & Athwal, R. S. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8872–8876 (1989).
8. Ishizaki, K. *et al.* *Mutat. Res.* **235**, 209–215 (1990).
9. Searle, A. G. *et al.* *Ann. hum. Genet.* **53**, 89–140 (1989).
10. Sanger, F., Miklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
11. Kozak, M. *Nucleic Acids Res.* **15**, 8125–8132 (1987).
12. Chou, P. Y. & Fasman, G. D. *Adv. Enzymol.* **47**, 45–148 (1978).
13. Evans, R. M. & Hollenberg, S. M. *Cell* **52**, 1–3 (1988).
14. Husain, I. *et al.* *J. biol. Chem.* **261**, 4895–4901 (1986).
15. Uchida, K. *et al.* *Biochem. biophys. Res. Commun.* **148**, 617–622 (1987).
16. Cherney, B. W. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 8370–8374 (1987).
17. Seeberg, E. & Steinum, A. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 988–992 (1982).
18. Friedberg, E. C. *DNA Repair* 213–264 (Freeman, New York, 1985).
19. Takebe, H. *et al.* *Cancer Res.* **37**, 490–495 (1977).
20. Yoshida, M. C. *et al.* *Jap. J. Cancer Res.* **77**, 1059–1061 (1986).
21. Committee on Standardized Genetic Nomenclature for Mice: Standard karyotype of the mouse *J. Hered.* **63**, 69–73 (1972).
22. Nesbitt, M. N. & Francke, U. *Chromosoma* **41**, 145–158 (1973).
23. Moran, E. & Mathews, M. B. *Cell* **48**, 177–178 (1987).
24. Call, K. M. *et al.* *Cell* **60**, 509–520 (1990).
25. Sukhatme, V. P. *et al.* *Cell* **53**, 37–43 (1988).
26. Kadonaga, J. T. *et al.* *Cell* **51**, 1079–1090 (1987).

ACKNOWLEDGEMENTS. We thank A. Asano for sequence analysis of the XPAC protein, and K. Matsubara for chromosome mapping. This work was supported by the Foundation for Promotion of Cancer Research and the Ministry of Education, Science and Culture of Japan.

Cloning of a mitogen-inducible gene encoding a κ B DNA-binding protein with homology to the *rel* oncogene and to cell-cycle motifs

Vincent Bours*, Juanita Villalobos*, Parris R. Burd*, Kathleen Kelly† & Ulrich Siebenlist*‡

*Laboratory of Immunoregulation, National Institutes of Allergy and Infectious Diseases, and †Laboratory of Pathology, National Cancer Institute, The National Institutes of Health, Bethesda, Maryland 20892, USA

WE have cloned and characterized a mitogen-inducible gene isolated from human T cells that predicts a protein of 968 amino acids. The amino-terminal domain has regions homologous to the oncogene *rel* and to the developmentally important gene *dorsal* of *Drosophila*. The carboxy-terminal domain contains repeat structures found in a variety of proteins that are involved in cell-cycle control of yeast and in tissue differentiation in *Drosophila* and *Ceanorhabditis elegans*, as well as in the putative human oncogene *bcl-3* and in the ankyrin protein. A truncated form of the product of this gene translated *in vitro* is a DNA-binding protein which interacts specifically with the κ B binding site found in many inducible genes, including the enhancer in human immunodeficiency virus. This gene is yet another in a growing list of important regulatory molecules whose expression is transcriptionally induced upon cellular activation.

Activation of resting cells with mitogenic agents initiates a sequential and interrelated programme of events which leads ultimately to cell proliferation and the expression of an activated phenotype. To gain insight into this programme, we used a subtractive strategy to clone in excess of 60 distinct genes that are induced early by mitogenic or antigenic stimulation of human peripheral blood T lymphocytes^{1,2}. Several transcription factors and/or nuclear oncoproteins are known to be encoded by genes that are induced early, including members of the *jun* and *fos* family^{3,4} and *c-myc*⁵. Among the novel inducible genes we have characterized are three encoding putative transcription factors: two zinc finger-containing DNA-binding proteins, designated 225 (ref. 6) and 591 (our unpublished observation; identical to early growth response gene-2 (ref. 7)), and gene 416, a member of the steroid-receptor family (our unpublished

Received 12 July; accepted 15 August 1990.

1. Tanaka, K. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 5512–5516 (1989).
2. Chen, C. & Okayama, H. *Molec. cell. Biol.* **7**, 2745–2752 (1987).

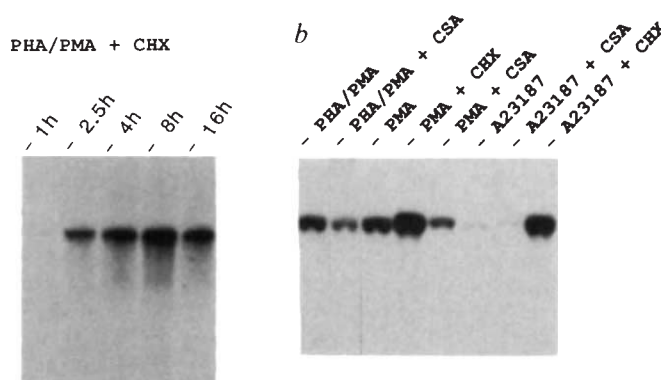
‡ To whom correspondence should be addressed.

FIG. 1 Induction of gene 243. *a*, Peripheral blood T lymphocytes were stimulated with phytohaemagglutinin (PHA; $1 \mu\text{g ml}^{-1}$) and phorbol 12-myristate 13-acetate (PMA; 25 ng ml^{-1}) in the absence or presence of cycloheximide (CHX, $10 \mu\text{g ml}^{-1}$). Total cellular RNA was extracted at the times indicated. *b*, Peripheral blood T lymphocytes were stimulated for 3 hours with PHA/PMA ($1 \mu\text{g ml}^{-1}$ and 50 ng ml^{-1} , respectively), PMA alone (10 ng ml^{-1}) and the calcium ionophore A23187 ($0.2 \mu\text{M}$), with or without CHX ($10 \mu\text{g ml}^{-1}$) or cyclosporin A (CSA, $0.5 \mu\text{g ml}^{-1}$), as indicated.

METHODS. Purification of peripheral blood T cells, extraction of RNA, fractionation on agarose/formaldehyde gels and hybridization conditions were as described^{1,2}. The hybridization probe lay between nucleotides 200 and 810 (see Fig. 2). Roughly equivalent

observation). We report here the sequence and characterization of the complementary DNA from gene 243, which encodes a DNA-binding factor that interacts specifically with κB sites.

Messenger RNA from 243 appeared first between 1 and 2.5 hours after stimulation of peripheral blood T cells with phytohaemagglutinin (PHA) and phorbol-12-myristate 13-acetate (PMA) and declined later, at about 16 hours (Fig. 1*a*). The combination of cycloheximide and mitogens superinduced 243 mRNA, suggesting that prior protein synthesis was not required for induction. A small amount of mRNA was also detected in unstimulated cells. PMA alone was a potent activator for 243 mRNA, whereas the calcium ionophore A23187 was a weak inducer in T cells. Addition of cycloheximide strongly augmented both signals (Fig. 1*b*). Cyclosporin A partially inhibited the activation mediated by PHA/PMA or PMA alone



amounts of RNA were loaded in each lane, as confirmed by hybridization with β_2 microglobulin (data not shown), except for the last lane in *b*, which contains about three times as much RNA.

in peripheral blood T cells (Fig. 1*b*), as well as by PHA/PMA but not PMA alone in the Jurkat T-cell line^{1,8}. Gene 243 is inducible in primary embryonic human lung fibroblasts⁸; it is also induced in human *Staphylococcus aureus* strain Cowan PMA-activated peripheral blood B cells (data not shown) and is transcriptionally activated, as established by nuclear run-on analysis in Jurkat T cells and lung fibroblasts⁸.

The cDNA sequence obtained from 243 is 3,838 nucleotides long, which is close to the estimated mRNA size of 4.3 kilobases (Fig. 2). The predicted open reading frame of 968 amino acids starts with a methionine codon at nucleotide 210, which conforms to the Kozak rules for initiator methionines⁹. The N-terminal region from amino acids 35–365 is strongly related to *v-rel*¹⁰, to *c-rel* from turkey¹⁰, chicken¹¹, mouse¹² and human¹³, and to the *Drosophila* gene *dorsal*¹⁴ (Fig. 3*a*). The gene *v-rel* is

```

..AATTCGGGCGCGGCTGCGGCGCATGCGCGGGCGCTGACTGGCTGGCGCGGCGCGCGCGGCTCCCGCTCGC 72
CCGACCGCCACTCGGCGCGCGCGCGGCTCGGCGCTGCGCGCGCTCTTCTCTCCAGCGCGCAGCGCGCG 144
CGGCTTAGAGGGAGAGCGCCACCGCGCGCAGAGCGCGCAACGCGGACTCGGCAACCGCGCTTCAGATGGCGAG 216
2
AAGATGATCCATTTGGGAGGCGTGAACAAATGTTTCATTTGGATCTCTTTGATCATACAAATATTTA 288
E Q Q P F E H L D F S L T H T I F
AAGAGAGATTTCAACACAGATGGCATGCCAACAGATACCTTCAATATTAGAGCAACCTA 360
N P E V F T Q F C M V D G G F Y L Q I L H V K Q F
1 -> REL Homology Start
AACAGAGGATTTGGTTTCGTTATGATGTAAGGCCATCCCATGGTGACTACCTGGTGCCTCTAGTG 432
K F R F E R Y V C E G P S H G G L P G G A S S
AAGAGAGAGATTTCAACCTCAGGTCAAAATCTGCAATATGTTGGACAGCAAGGTTATTTCTCAGT 504
E K N K K S Y P Q V K I C N Y V G P A K V I V Q
TGGTCAAGATGAAATATCCCATGTCATGCCACAGCGCTGGTGGGAAACACTGTGAGGATGGGATCT 576
L V T N G K N I H L H A H S L V G K H C E D G I
GCACTTAAGTCTGGAGCCAGGACGATGGTGGCTTCGCAAACTGGGTATACCTCATGTGACAAAGA 648
C T V T A G P K D M V V G F A N L G I L H V T K
AAAAATTTTGAAGACTGGAGCAAGCAAGTGCAGAGCGGTATATAAGGGCTATATCTGGACTCTTGG 720
K K V F T E A C L E A R M T E A C I R G Y N P G L
TGCACCTGACCTTGGCTATTTGCAAGGCAAGTGGAGGGGACCGGACGCTGGGAGATCGGGAAGAGC 792
V H F D L A Y L Q A E G G G D R C Q L G D R E K E
TAATCCGCAAGCAGCTTGCAGAGCAGCAAGAGATGACCTCAGCGTGGTGGGCTCATGTTTACAGCT 864
L I R Q A A L Q Q T K E M D L S V G T L R M F T A
TTCTTCGATAGCATGCGCATGCAAGAGCGCTTCAAGAGCGGTATGATGAGCGCATCTATGACAGTA 936
F L F D S T G S T G F R R L E P V V S D A I Y D S
AAGCGCCATGATGATCAACTTGAATTTGAAGATGACAGGACAGCTGGATGCTGATGAGGAGGAGG 1008
K A F P S N L K I V R M D R T A G C T V T G G E
AAATTTATCTCTTTGTGACAAAGTTGACAAAGATGACATGACATGATTTTGAAGAGGAGAAATG 1080
E I Y L L C D K V Q K D D I Q R F I Y E E E N
GTGGAGTCTGGGAGGATTTGGAGATTTTTCGCCACAGATGTCATAGACAAATTTGCCATTTGCTCTCAA 1152
G G V W E G F P D S F D V H R Q F A I V F K S
CTCCAAAGTATAAGATATTAATTTACAAAACAGCGCTCTGTTTGTCCAGCTTCGGAGGAAATCTGACT 1224
T F E Y K D Q N I T K F A S V F V F L R R R S D
TGCAAGTACTGAAACAAACCTTTCTCTACTATCTGAAATCAAGATAGAGAGAGTGCAGAGGAAC 1296
G E A C T F E F D S F E B K D W E E V Q R K
GTCAAGAGCTATGCCCAATTTTTCGATAGTTTTCGCGCGTGGTATGGTGGCGGAGCTGGAGGAGGCA 1368
R C A G A G T G G G G G G G G T G S T G T G P G Y S F P H Y G
1 -> Rel Homology End
1 -> Glycine-rich Start
TGTTTGGTATGGCGGCTGGAGGAGGCGCTGGAAGTACAGGTCAGGCTGCAAGCTTCCACATCTGGAT 1440
M F F G S G G G G G G G T G S T G T G P G Y S F P H Y G
Glycine-rich End
TTCTACTTATGTTGGGATTTACTTCCATCTGGAATCAATTAATGATGGATGAAGCATGAACCA 1512
F F Y Y G G I T F H F G T T K S N A G M K H G T
TGGACGATCAATTAAGAGGACCTTGAGGTTGTGACAAAGTATGACAAACACTGTAAACCTCTTGT 1584
M D A T E S K K K D F E G C D K S D D K N T V N L F
GGAAGTTTATTAAGACAGAGAGTACAGGAGCGGACGAGGACACGCTGGGAATGTTGAGGTCACTC 1656
G K V I E T T E Q D Q E F S E A T V G N G E V T
TAACCTATGCAACAGGAAACAAAGAGAGTGGTGAAGTCAGGATCACTCTGTTCTGACAGAGGCTATGC 1728
L A T Y A T G T K E E S A G V Q D N L F L E K A M
506

```

FIG. 2 Nucleotide sequence and predicted amino-acid sequence of 243. The *rel*-related domain, the glycine-rich stretch and the repeat motifs are indicated. The individual repeats (including two half-repeats) are underlined, so is the poly(A)-addition site. Numbering of nucleotides and amino acids is shown on the right. *EcoRI* cloning sites are included. In addition to the

```

AGCTTGCAGAGGAGCATGCCAATGCCCTTTTTCGACTACCGGCTGACAGGAGCTGAAGATGCTGCTGGCG 1800
Q L A K R H A N A L F D Y V T S D I K M L L L A
Repeats Start 1 ->
TCCAGCGCATCTCACTGCTGCGAGGATGAGAAATGGGACAGCTGTACACTTACCACTATCCACCTTC 1872
V Q R H L T A V Q D E N G D S V L H L A I I H L
ATTCTCAACTTGTAGGGATCTACTAGAAATGACATCTGGTTTGAATTTCTGATGACATTCACACATGAGAA 1944
H S Q L V R D L L E V T S G L I S D D I I N M B
ATGATCTGTACAGAGCGCTTGCATCTGGCACTAAGCAGGAAGATGTGGTGGAGGATTTGTCTGA 2016
N D L Y Q T F L H L A V I T K Q E D V V E D L L
GGGCTGGGCGGCACTGAGCGCTTGGAGCGCTTGGTAATCTGTTTGGACATGAGTGCCAAAGAGAGC 2088
R A G A D T L S L L D R L G N S V L L H L A A K E G
ATGATGAATTTCTCAGTATCTTACTAAGCAAAAAGGACAGCACTTCTTGACCCACCCCAACGGGAGC 2160
H D K V L S I L L K H K K A A L L L D H P N G D
GTCTGAATGCCATTCATCTAGCATGATGAGCAATGACCTGCCATGTTTGTCTGCTGCTGGGCGCTGGG 2232
G L N A I H L A M M S N S L P C L L L L V A A G
CTGACGTCATGCTGAGGAGGAGTCCGCGGCGGACAGCACTGCACCTGGCTGTGGAGCAGCAACATCT 2304
A D V N A Q E Q K S G R T A L H L A V E H D N I
CATTGGCAGGCTGCTGCTCTGGAGGATGCCATGCTGGACAGTATCTACGATGGAACCAACCC 2376
S L A G C L L L L E G D A R V D S T T Y D G T T P
TGCAATATGAGCTGGAGAGGCTCAGCAGGCTGGACCTTCTCAAGCAGCAGGAGCATGCTCCCTGG 2448
L H I A A G R G S T R L A L A A G A D P L
TGGAGATCTTGAAGCTCTTATGACCTGGATGCTCTGGGAAAGTACAGGAGGATGAGGAGGATTTGTC 2520
V E N F E F L Y D L D D S W E N A G E D E G V V
CTGGAACCGCTGTAGATGGCAGCAGCTGGCAGGATTTGACATATTAATGGGAACCATATGAGC 2592
E G T T F L D M A T S W Q V F D I L N G K P Y E
1 -> Repeats End
CAGATGTTACATCTGATGATTTAGTACAGAGGAGACATGAACAGCTGCGCTGAGATGTGAAGCTCAGC 2664
P E F T S D D L L A Q G D M K Q C L A E D V K L G
TGATATGATAGTAAATCTGATGACAGCAAAATGCTGGCTACTTGGCGCAGAAATAGCTGTCCGGA 2736
L Y K L L E I P D P D K N W A T T G A Q K L G L G
TACTTAAATGCTTCCGCGTGAAGTCTGCTCTTCCAAACACTATGACCAATGAGGCTCTGTGGG 2808
I L N N A P F R L S P A P K C T L M D N Y E V S G
GTACAGTCAGAGAGCTGTGGAGGCGCTGAGCAAAATGGGCTACACGCAAGCAATGAAGTATGACCA 2880
G T V R E L V E A L R L Q M G Y T E A I E V I Q A
CCTCCAGCCAGTGAAGACACCTCTCAGGCGCACTGCTGCTCTGCTGCTGCTCCAGGAGGAGCA 2952
A S S P V K T T S Q A H S L L P L S P A S T R Q Q
TAGACGCTGCGAGCAGCTGCTGCTGCGAGCGGCTGGAGACATCTCTGCGCAACCTACGTTTA 3024
I D E L R D S D S V C D S G V E T S T F G R K L S F
CCGAGTCTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3096
T E S L T S G A S L L T L N K M P H D Y G Q E G
CTCTAAGGCAAAATTTAGCTGCTGACATTTCCCAACCGGTGAACCAAGGCTCAAAATTTCCATCT 3168
P L E G K I
GTGTCACAGAGCAGAGCTGAAGTGCATCCAAAGGTGCTCAGAGAGCGCGCGCGCTGAATCATCTCGA 3240
TTTAACCTCGAGACCTTTCACTTGGCTTCTCTTCTGTTGTTCAAAATGAATTTAGTTGTTGTTCACT 3312
GATAGTATCTAGCAATCACACACTGGCTGAGCGGATGCTGCTGGGATGAGGTTGCTTACTAAGCTT 3384
AGCTGCTGCTGATCAGCTGCTTCTGTTGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3456
AAGGATACCGCTGCGCACTGGCATCTTCTGACCAAGCATCATTTTGCATCAAAATTAAGGCTTAAGA 3528
AAAGAGATATTTAAATGAGAGTCACCTGATGTGCATTTTAAAGAAAGGAGCATTTGCTTTTCTAATG 3600
TGGTTATTTCTGATTTGCAAAAAAAGAAAAAATCTTGTCAATATTTAAACATGTTTAC 3672
AATCATTTGAAAAAGGTTATTTTCCCTGTTTCTGATTTTCTGATTTTCTGATTTTCTGATTTTCTG 3744
ACTTTAAAGGAAAAATTTGGAATTTAATGCTATTTTATTTTACTTTTAAATAAAGGAAAAAGCA 3816
TTGATGACCTCAGGAATTC 3838

```

presumed translational start site, there are two in-frame ATGs at amino-acid positions 14 and 35; there are several in-frame stop codons in the G+C-rich untranslated leader sequence. Nucleotide sequencing of multiple overlapping cDNA clones was by the dideoxy chain termination method (Sequenase, US Biochemical), following the manufacturer's instructions.

the transforming component of the replication-deficient virus REV-T, which is extremely oncogenic¹⁰. The *Drosophila* maternal-effect gene *dorsal* encodes the essential morphogen that determines ventral-dorsal polarity in developing embryos; it regulates genes by entering ventral but not dorsal nuclei from the cytoplasm 90 min after fertilization¹⁵⁻¹⁷. Both *dorsal* and *rel* genes in tissue culture cells transactivate various promoters by means of their N-terminal domains^{15,18,19}.

Besides this homology with *rel*-related proteins, protein 243 contains repeat structures in its C-terminal half which are phylogenetically conserved and are evident in other genes, ranging from those controlling the yeast cell cycle to the human gene for ankyrin, which links integral membrane proteins to the cytoskeleton (see Fig. 3b for examples). There are six full repeats of 33 amino acids and two half-repeats in protein 243 (Fig. 3b). Although no common function has been attributed to these repeats, the roles of some of the genes containing them suggest an intriguing association with the cell cycle, with signal transduction and with the cytoskeleton. These repeats could form helix-loop-helix structures in which the helices would be highly amphipathic and therefore good candidates to determine protein-protein interactions, possibly with cytoskeletal-like structures²⁰.

The *dorsal*¹⁵⁻¹⁷ and *rel*²¹ gene products resemble NF- κ B, a nuclear DNA-binding complex essential for expression of many inducible genes²². The NF- κ B complex consists of two subunits, p50 and p65 (proteins of relative molecular masses 50,000 and 65,000 (50 K and 65 K) respectively), of which p50 binds DNA directly²³. NF- κ B is normally complexed in the cytoplasm with

its inhibitor I- κ B, from which it is released upon cellular activation to be translocated into the nucleus^{23,24}. To test whether the 243 protein could interact with the κ B sequence, we first assayed for transcription and translation *in vitro* using DNA templates truncated at different sites starting from the N-terminal coding sequence and ending beyond the C-terminus (Fig. 4b). This gave sequentially longer proteins and confirmed that the predicted open reading frame was correct (data not shown). We used these synthesized proteins in an electrophoretic mobility shift assay with a labelled NF- κ B-binding oligonucleotide derived from the human immunodeficiency virus enhancer (Fig. 4a). The truncated protein resulting from the *Xba*I digestion (lane C) bound efficiently, but longer proteins gave little or no binding (lanes D-F), and shorter proteins did not bind at all (lanes A, B). This indicated that sequences homologous to *rel* protein are necessary for binding and that the C-terminal domain of 243 protein contains a sequence that inhibits binding.

To investigate the specificity of binding to the κ B sequence, we assayed for competition by electrophoretic mobility shift analysis. The protein from the *Xba*I truncation was incubated with HIV- κ B oligonucleotides or with κ B oligonucleotides derived from the functional site of the interleukin-2 (IL-2) gene promoter, and competed with an excess of non-labelled mutant or wild-type oligonucleotides (Fig. 4a, right panel). The mutants did not compete in either case, whereas wild-type oligonucleotides eliminated all binding. Also, a labelled mutant IL-2 oligonucleotide did not produce a shift.

Our results indicate that 243 protein is an NF- κ B-like protein which we designate NF- κ B for transcriptionally activated bind-

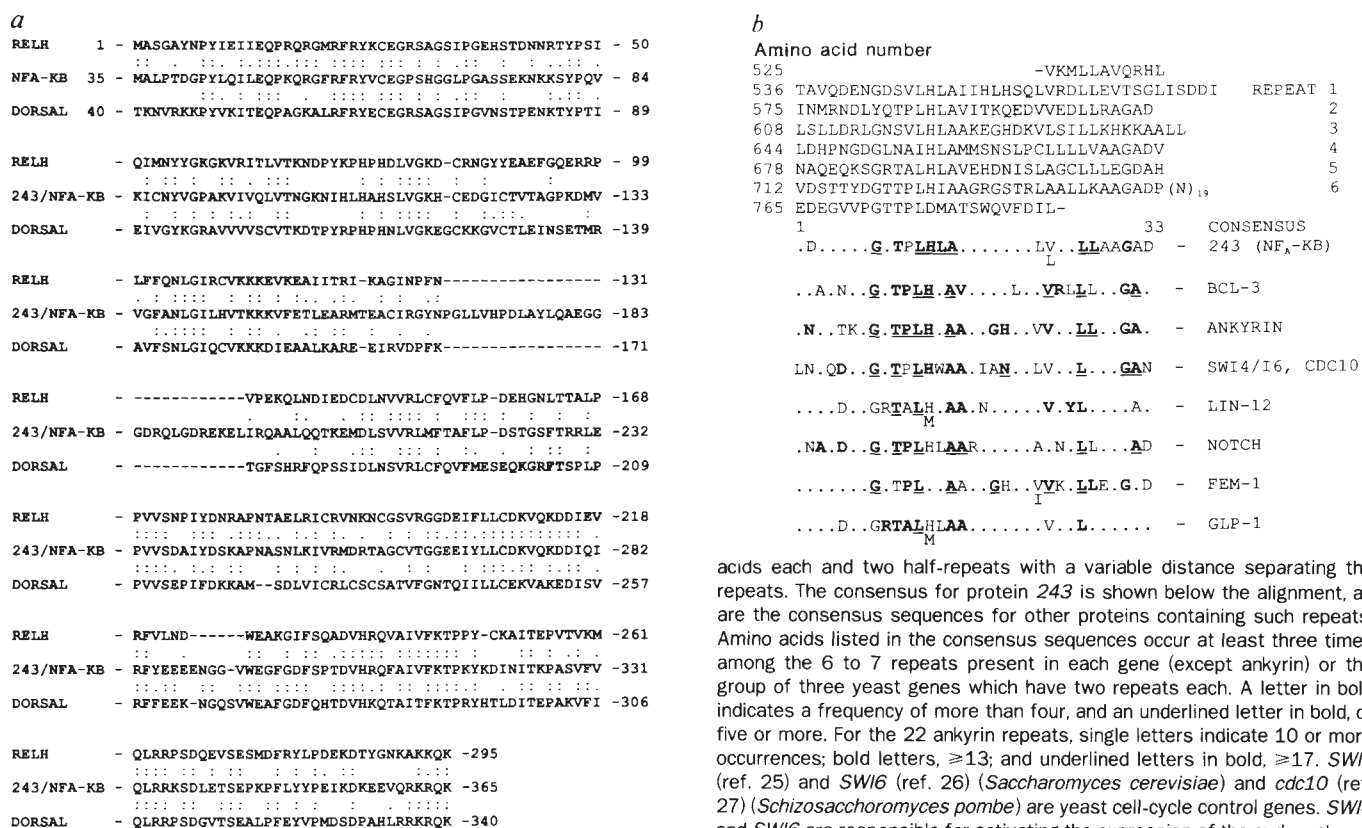


FIG. 3 *a*, Amino-acid comparison of the N-terminal domains of 243, human *c-rel* (RELH) and *dorsal* proteins. Double dots indicate sequence identity and single dots denote similar amino acids (in one-letter notation: A,S,T; D,E; N,Q; R,K; I,L,M,V; F,Y,W). Proteins 243 and *c-rel* are identical in 48.8% of their residues (over a length of 295 amino acids), and 243 and *dorsal* are identical in 45.5% of their amino acids (over 301 amino acids). *Dorsal* and *c-rel* proteins have 48.1% of their amino acids identical (data not shown). *b*, Consensus amino acids for the cell-cycle repeats. Sequentially numbered amino acids of the 243 gene are aligned for maximum homology with the cell-cycle repeats underlined in Fig. 2. There are six full repeats of 33 amino

acids each and two half-repeats with a variable distance separating the repeats. The consensus for protein 243 is shown below the alignment, as are the consensus sequences for other proteins containing such repeats. Amino acids listed in the consensus sequences occur at least three times among the 6 to 7 repeats present in each gene (except ankyrin) or the group of three yeast genes which have two repeats each. A letter in bold indicates a frequency of more than four, and an underlined letter in bold, of five or more. For the 22 ankyrin repeats, single letters indicate 10 or more occurrences; bold letters, ≥ 13 ; and underlined letters in bold, ≥ 17 . *SWI4* (ref. 25) and *SWI6* (ref. 26) (*Saccharomyces cerevisiae*) and *cdc10* (ref. 27) (*Schizosaccharomyces pombe*) are yeast cell-cycle control genes. *SWI4* and *SWI6* are responsible for activating the expression of the endonuclease HO during a short phase of the yeast cell cycle in late G1; *cdc10* is essential for progression of the cell cycle through G1. The *Drosophila* neurogenic locus *notch* (28) and the *C. elegans* genes *lin-12*, *glp-1* (ref. 29) and *fem-1* (ref. 30) (each with 6 repeats), all determine cell fates during development; they each encode integral transmembrane proteins, with the exception of *fem-1*, which encodes a cytoplasmic protein. The gene *bcl-3* (ref. 31) is a putative human oncogene (7 repeats). The erythrocyte ankyrin repeat domain (22 contiguous repeats) functions to bind the erythroid anion exchanger and tubulin²⁰. Not shown are repeats present in host-range proteins of vaccinia, cowpox and fowlpox virus²⁰.

representing the two κ B binding sites of the human immunodeficiency virus (HIV) enhancer. Lane C shows strong binding with a protein which includes all of the homologous regions of *rel* but excludes the repeat motif (*Xba*I cut in *b*). Much of the residual binding with the longer proteins (lanes E, F) could have resulted from degradation or incomplete synthesis in the wheat germ system to produce small amounts of proteins similar in size to that from the *Xba*I truncation, an interpretation supported by the migration of all the various shifts to roughly the same position, even though the proteins should be increasingly larger. The *Xba*I-truncated protein was used for the competition experiments shown in the right panel. Binding to the HIV- κ B was competed with the wild-type oligonucleotide (wt), or a mutant (m)

Similarly, binding to a κ B site derived from the IL-2 promoter³⁴.

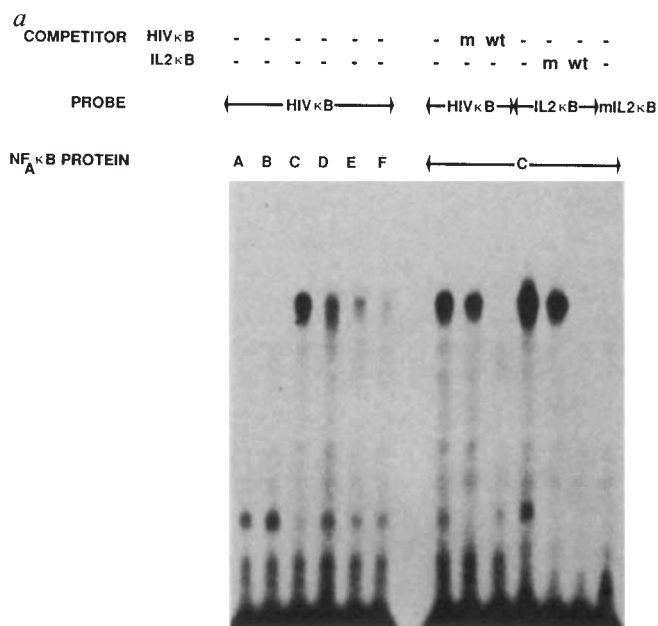
was competed with the wild-type, or with a mutant oligonucleotide

The last lane in the right panel demonstrates that a labelled mutant IL-2 κ B site does not produce a shift. *b*, *In vitro*-synthesized proteins are represented by a line and marked A-F. The restriction enzymes used to cut the template DNA were: A, *SphI* at nucleotide position 533; B, *PstI* at 810; C, *XbaI* at 1,711; D, *Scal* at 2,349; and E, *SacI* at 2,958 (see Fig. 2). To produce the full-length protein, the DNA was cut with *DraI* at position 3,539. Rel-related and repeat domains are boxed. Numbers refer to amino-acid positions (see Fig. 2).

METHODS. Linearized vector (2 µg; Bluescript) plus insert was transcribed into RNA using T7 polymerase (Stratagene) as recommended by the manufacturer. The DNA template was removed by DNase I digestion. Following phenol and chloroform extraction and ethanol precipitation, one third of the RNA was used for *in vitro* translation with the wheat germ extract, as recommended by the manufacturer (Promega). Six per cent of the translated product was incubated for 30 min with 0.2 ng labelled oligonucleotides

ing protein. Our data do not address the question of whether this gene encodes the p50 DNA-binding subunit of the NF- κ B complex²³, although the experiment shown in Fig. 4 suggests that 243 protein, with its calculated molecular weight of 105K, may be processed *in vivo* to a protein of ~50K, which is similar in size to that resulting from the *Xba*I truncation. The putative proteolytic cleavage site of 243/NF- κ B would then fall at or near a glycine-rich stretch of amino acids (Fig. 2). There may be additional κ B-binding proteins: *rel* and *dorsal* proteins may find κ B as well. We are also currently characterizing another *rel*-related gene present among our induced clones.

Although NF- κ B is preformed in cells and can be activated in the presence of protein synthesis inhibitors, this does not address that NF $_{\Lambda}$ - κ B, a possible component of this complex, is transcriptionally induced. κ B-binding activity may be regulated at several levels. Considering the roles possible for the cell-cycle motifs, we speculate that the newly induced NF $_{\Lambda}$ - κ B protein could be sequestered in the cell by attachment to cytoskeletal-like structures. A subsequent cellular signal, possibly associated with the cell cycle, may cause the release and/or processing of the factor to a DNA-binding form. Such a control mechanism would ensure that genes regulated by NF $_{\Lambda}$ - κ B are only induced transiently, the amount of processed protein being limited until replenished. Furthermore, such a mechanism would allow for linkage to the cell cycle. It may be important in T cells, for example, to coordinate the synthesis of DNA during the cell-cycle with the induction of genes such as lymphokines.



b

(100,000 c.p.m.) in incubation buffer (Stratagene). A 200-fold excess of unlabelled oligonucleotides were used for the competition assays. The resulting complexes were separated on 4% polyacrylamide gel in the cold in PGE buffer (50 mM Tris-HCl, 0.38 M glycine, 2 mM EDTA, pH 8.5). The oligonucleotides were labelled by filling in the overhangs with Klenow enzyme in the presence of radioactive nucleoside triphosphates.

Note added in proof: After submission of this manuscript two reports appeared of the cloning of the human³⁵ and mouse³⁶ p50 subunit of NF- κ B. The human gene for p50 is identical to NF- κ B.

Received 28 August; accepted 24 September 1990.

1. Zipfel, P. F., Irving, S. G., Kelly, K. & Siebenlist, U. *Molec. cell. Biol.* **9**, 1041-1048 (1989).
2. Irving, S. G., June, C. H., Zipfel, P. F., Siebenlist, U. & Kelly, K. *Molec. cell. Biol.* **9**, 1034-1040 (1989).
3. Ryder, K. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8464-8467 (1988).
4. Zerial, M. *et al.* *EMBO J* **8**, 805-813 (1989).
5. Kelly, K., Cochran, B. H., Stiles, C. D. & Leder, P. *Cell* **35**, 603-610 (1983).
6. Wright, J. J. *et al.* *Science* **248**, 588-591 (1990).
7. Joseph, L. J. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 7164-7168 (1988).
8. Gunter, K. C., Irving, S. G., Zipfel, P. F., Siebenlist, U. & Kelly, K. *J. Immun.* **142**, 3286-3291 (1989).
9. Kozak, M. *J. cell. Biol.* **108**, 229-241 (1989).
10. Wilhelsen, K. C., Eggleton, K. & Temin, H. M. *J. Virol.* **52**, 172-182 (1984).
11. Chen, I. S. Y., Wilhelsen, K. C. & Temin, H. M. *J. Virol.* **45**, 104-113 (1983).
12. Grumont, R. J. & Gerondakis, S. *Oncogene Res.* **4**, 1-8 (1989).
13. Brownell, E., Mittereder, N. & Rice, N. R. *Oncogene* **4**, 935-942 (1989).
14. Steward, R. *Science* **238**, 692-694 (1987).
15. Rushlow, C. A., Han, K., Manley, J. L. & Levine, M. *Cell* **59**, 1165-1177 (1989).
16. Steward, R. *Cell* **59**, 1179-1188 (1989).
17. Roth, S., Stein, D. & Nusslein-Volhard, C. *Cell* **59**, 1189-1202 (1989).
18. Hannink, M. & Temin, H. M. *Molec. cell. Biol.* **9**, 4323-4336 (1989).
19. Kamens, J., Richardson, P., Mosialos, G., Brent, R. & Gilmore, T. *Molec. cell. Biol.* **10**, 2840-2847 (1990).
20. Lux, S. E., John, K. M. & Bennett, V. *Nature* **344**, 36-42 (1990).
21. Davis, N., Bargmann, W., Lim, M.-Y. & Bose Jr, H. R. *J. Virol.* **64**, 584-591 (1990).
22. Lenardo, M. J. & Baltimore, D. *Cell* **58**, 227-229 (1989).
23. Baeuerle, P. A. & Baltimore, D. *Genes Dev.* **3**, 1689-1698 (1989).
24. Ghosh, S. & Baltimore, D. *Nature* **344**, 678-682 (1990).
25. Andrews, B. J. & Herskowitz, I. *Nature* **342**, 830-833 (1989).
26. Breeden, L. & Nasmyth, K. *Nature* **329**, 651-654 (1987).
27. Aves, S. J., Durkacz, B. W., Carr, A. & Nurse, P. *EMBO J.* **4**, 457-463 (1985).

28. Artavanis-Tsakonas, S. *Trends Genet.* **4**, 95-100 (1988).
29. Yochem, J. & Greenwald, I. *Cell* **58**, 553-563 (1989).
30. Spence, A. M., Coulson, A. & Hodgkin, J. *Cell* **60**, 981-990 (1990).
31. Ohno, H., Takimoto, G. & McKeithan, T. W. *Cell* **60**, 991-997 (1990).
32. Singh, H., Sen, R., Baltimore, D. & Sharp, P. *Nature* **319**, 154-158 (1986).
33. Nabel, G. & Baltimore, D. *Nature* **326**, 711-713 (1987).
34. Hoyos, B., Ballard, D. W., Bohnlein, E., Siekevitz, M. & Greene, W. C. *Science* **244**, 457-460 (1989).
35. Kieran, M. *et al.* *Cell* **62**, 1007-1018 (1990).
36. Ghosh, S. *et al.* *Cell* **62**, 1019-1029 (1990).

ACKNOWLEDGEMENTS. We thank A. S. Fauci for support, encouragement and critical reading of the manuscript, S. Park for technical assistance, and J. Kehrl for B-cell RNAs. V.B. was supported by fellowships from the Belgian Work Against Cancer and the Belgian National Fund For Scientific Research.

Phosphorylation of the C terminus of Fos protein is required for transcriptional transrepression of the *c-fos* promoter

Rivka Ofir, V. J. Dwarki, Dana Rashid & Inder M. Verma*

Molecular Biology and Virology Laboratory, The Salk Institute, PO Box 85800, San Diego, California 92138, USA

PROTO-ONCOGENE *fos* encodes a nuclear phosphoprotein of 380 amino acids that can modulate the transcription of other genes either by transactivation or by transrepression¹⁻⁷. The v-Fos protein (381 amino acids) shares the first 332 amino acids with the c-Fos protein (with five single amino-acid changes), but differs at the C terminus⁸. We have previously reported that the c-Fos protein undergoes more extensive post-translational modification than v-Fos (refs 9, 10). The major modification of the c-Fos protein involves serine phosphoesterification of sites in the extreme C terminus¹⁰. We therefore argued that modification of the C-terminal region of the c-Fos protein may be involved in its ability to transrepress transcription without compromising its ability to transactivate other genes. Here we show that mutant c-Fos protein which is hypophosphorylated at its C terminus is unable to repress transcription of the *c-fos* promoter following induction with serum or tetraphorbol acetate. The C-terminal phosphorylation-deficient mutant is, however, fully competent to activate transcription of promoters containing a phorbol response element. The requirement for phosphorylation can be offset by the introduction of a net negative charge in the C terminus of the Fos protein.

A diagrammatic sketch of the various domains of the Fos protein with the amino-acid sequence of the C-terminal domain is shown in Fig. 1a. The C-terminal serine residues are found in three groups (Ser 362-364; Ser 368-369 and 371; Ser 373-374). The amino-acid sequence RKGSSS (single-letter notation) (359-364; Fig. 1) is an excellent consensus recognition site for cyclic-AMP-dependent protein kinase¹¹. As any one of the serine residues between amino acids 362-374 was a likely site of phosphoesterification, we mutated groups of serine residues together to alanines to give the following mutants: *fosSerA*, serine residues at position 362-364 mutated; *fosSerB*, 368, 369 and 371 mutated; and *fosSerC*, 373 and 374 mutated. All the mutated constructs were transcribed and translated *in vitro* and migrated on SDS-PAGE in accordance with their predicted molecular weights (data not shown). The constructs expressing mutant Fos proteins were tested for their ability to suppress transcription from FC4 (-404 to +42 of the *c-fos* promoter linked to marker chloramphenicol acetyltransferase (CAT) sequences). Percentage CAT conversion (CAT activity) in the presence of a wild-type *fos* construct was at least 10-fold lower than that observed with serum induction alone (Fig. 1b, compare

lanes 2 and 3). The level of CAT activity in lane 3 was equivalent to that observed in starved cells (lane 1), demonstrating that the wild-type Fos protein can repress transcription from the *c-fos* promoter⁶. When constructs encoding FosSerA were cotransfected with FC4, no repression of CAT activity was observed (compare lanes 2 and 4). The level of *fos* promoter activity was similar to that following serum stimulation. Mutants *fosSerB* and *C* were also much less effective in repressing transcription from FC4 (lanes 5 and 8). A mutant Fos protein where the C-terminal 118 amino acids were deleted (FosΔ262, lane 7) was also unable to repress transcription of *fos* promoter. The viral Fos proteins (FBJ-MSV, Finkel Biskis Jinkins murine osteosarcoma virus, lane 8; or FBR-MSV, F. B. Reilly MSV, lane 9) which have altered C termini, do not repress transcription of the *fos* promoter. To rule out the possibility that any alteration in the C terminus would be detrimental to the suppressor activity of Fos protein, we engineered point mutations that altered Ala 356 and 357 to valine (Fos A356, 357V). This mutant was still

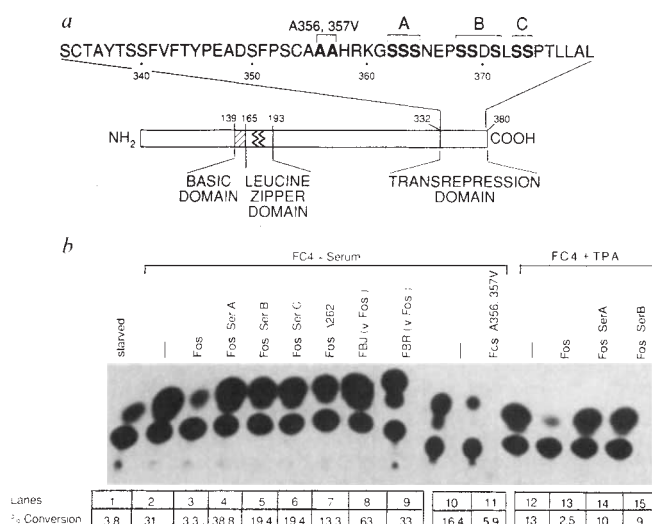


FIG. 1 Transrepression properties of *c-fos* and *fos* mutants. *a*, Diagrammatic sketch of the c-Fos protein; the positions of the mutations and other structural landmarks are indicated. *b*, Expression vectors (6 µg of each) generating c-Fos, v-Fos and c-Fos mutant proteins were cotransfected with FC4 (4 µg), and CAT activity was determined after serum stimulation (lanes 1-7) or TPA induction (100 ng ml⁻¹, lanes 8-10) of quiescent NIH 3T3 cells. The results are presented as per cent of conversion of the non-acetylated to acetylated form of chloramphenicol. Each sample was independently assayed for transrepression activity at least three times. METHODS. NIH 3T3 cells were transfected by the calcium-phosphate coprecipitation technique and manipulated as described⁶. CAT activity was determined as described¹⁴. All samples for CAT assay were standardized for protein content. Oligonucleotide-directed mutagenesis generated site-specific mutations in the C terminus of *c-fos* (ref. 15). A 300-base pair (bp) *NcoI-SalI* fragment from mouse *c-fos* was cloned into the vector M13 MP19. Oligodeoxynucleotides complementary to the region containing the specific serines to be altered were synthesized. The oligos are:

```

1.223          1.252
fosSerA 5' GCTGCCACCGAAAGGGC GCG GCG GCG AACGAGCCCC 3'
              Ala Ala Ala

1.237          1.267
fosSerB 5' GAGCCC GCG GCG GAC GCG CTG 3'
              Ala Ala

1.258          1.283
fosSerC 5' CGACTCCCTG GCG GCA CCCACGCTG 3'
              Ala Ala

1.210          1.239
fosA356, 357V 5' CCCAGCTGTGCA GTC GTA CACCGCAAGGGC 3'
                  Val Val

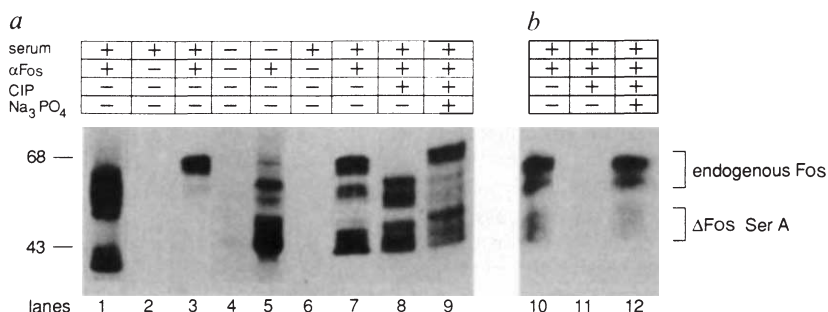
```

Other plasmids used have previously been described^{6,8,17-19}.

* To whom correspondence should be addressed.

FIG. 2 Metabolic labelling of c-Fos and Δ FosSerA proteins with ^{35}S and ^{32}P . *a*, NIH 3T3 cells were metabolically labelled with [^{35}S]methionine. Proteins were immunoprecipitated from denatured lysates except for lane 1 which shows both Fos and co-immunoprecipitated Jun proteins. Lane 2, NIH 3T3 cells with preimmune sera (— denotes preimmune sera); Lane 3, boiled lysates of NIH 3T3 cells with α Fos; lanes 4–9, boiled lysates from Δ fosSerA stable cell line; lane 4, with preimmune sera; lane 5, α Fos without serum stimulation; lane 6, serum stimulation with preimmune sera; lane 7, serum stimulation + α Fos; lane 8, serum stimulation + α Fos + CIP; lane 9, same as lane 8 + Na_3PO_4 . Autoradiograph was exposed for 16 h. *b*, The experiment was essentially as in *a*, but cells were metabolically labelled for 4 h, with [^{32}P]orthophosphate and induced with serum in the last hour of labelling. Boiled lysates were immunoprecipitated with α Fos (lanes 10–12). Lane 10, +serum; lane 11, +serum + CIP; lane 12, +serum + CIP + Na_3PO_4 . Autoradiograph was exposed for 16 h.

METHODS. Δ FosSerA was generated by three part cloning: BK vector (the expression vector BK-28 lacking the cDNA of c-fos), human c-fos lacking the region coding for amino acids 133–193 which extends to NcoI site and



able to repress transcription of the *fos* promoter (compare lanes 10 and 11), suggesting that phosphorylation of the C terminus of the Fos protein is essential for transrepression.

As tetraphorbol acetate (TPA) can induce *fos* transcription, we asked whether this induction is also repressed by Fos protein. Addition of wild-type *fos* plasmid decreases the TPA-induced CAT activity by at least fivefold (compare lanes 12 and 13). Once again, FosSerA and SerB mutant proteins were unable to repress the transcription from the *fos* promoter (lanes 14 and 15).

To confirm that phosphorylation of serine residues in the Fos C terminus is essential for suppression, we performed experiments to show that *in vivo* synthesized mutant Fos protein is hypophosphorylated as a result of mutations at the C terminus. We generated a shorter C-terminal mutant Fos protein which can be distinguished from the endogenous Fos protein on PAGE. Because the additional phosphorylation sites in c-Fos protein reside in the C-terminal 20 amino acids¹⁰, we constructed a *fos* plasmid with SerA mutations (Δ FosSerA; Fig. 1a) but lacking the region encoding amino acids 133–193, which would generate a truncated protein of 320 amino acids instead of the 380-amino acid wild-type protein. The truncated protein migrates on SDS-PAGE with an apparent relative molecular mass of ~43,000 (43K) (lane 5), which is easily distinguishable from the 55–68K wild-type Fos protein (Fig. 2a, lane 3). The Fos proteins were metabolically labelled with [^{35}S]methionine in NIH 3T3 cells stably transfected with the Δ FosSerA construct, followed by immunoprecipitation of the lysate with monoclonal anti-Fos

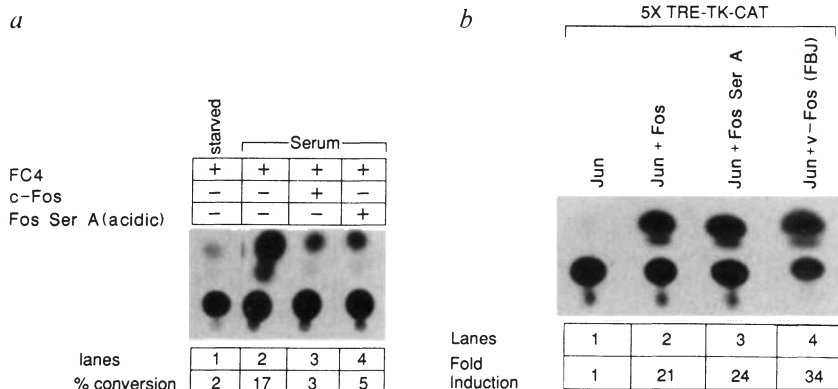
NcoI–EcoRI fragment from *fosSerA*. Cells from a stable cell line (10^6) containing Δ fosSerA were seeded in a 10-cm tissue culture dish 36–60 h before the experiment was performed. The cells were starved of serum for 24–48 h before labelling with [^{35}S]methionine, preincubated for 30 min and analysed as described^{10,12}. Cells were labelled with [^{32}P]phosphate by incubation in 3 ml DMEM with 0.5% dialysed FCS lacking phosphate for 3 h and analysed¹⁰. Immunoprecipitations were carried out with anti-*fos* 18H6 monoclonal ascites fluid¹² (typically 1 μ l per 10^6 cells) and proteins were resolved by 10% SDS-PAGE as described previously¹⁸.

antibody 18H6 (ref. 12). If the extracts are not boiled before immunoprecipitation, in addition to 55–68K Fos protein, the p39/Jun protein is also coimmunoprecipitated (Fig. 2a, lane 1). If, however, the lysate is boiled before immunoprecipitation, then the Fos–Jun complex is dissociated and only the Fos protein can be identified (Fig. 2a, lane 3). As the size of the shorter mutant protein is in the range 40–43K, we performed immunoprecipitation on denatured lysates. The truncated Fos–SerA mutant protein generated from a construct driven by a long-terminal repeat can be identified in unstimulated cells (lane 5). On stimulation with serum, the endogenous Fos protein migrating between 55–68K can also be identified (lane 7). When the extract is treated with calf intestinal phosphatase (CIP), the endogenous Fos protein migrates faster, but the exogenously introduced truncated FosSerA protein remains essentially unaffected (lane 8). The addition of phosphatase inhibitor shows that the effect on mobility shift is due to phosphoester cleavage (lane 9).

To confirm that the FosSerA mutant is hypophosphorylated at the C terminus *in vivo*, we also carried out parallel metabolic labelling with [^{32}P]orthophosphate, followed by immunoprecipitation of the Fos protein from denatured lysates. Figure 2b shows that the truncated FosSerA is poorly labelled compared with the endogenous Fos proteins (lane 10). Equimolar quantities of the protein were immunoprecipitated as judged by western blotting (data not shown). In contrast, the relative amounts of ^{35}S label in both the endogenous and the exogenous

FIG. 3 *a*, Repression by FosSerA-(acidic). Expression vectors (6 μ g) of c-Fos and FosSerA (acidic) were cotransfected with 4 μ g of FC4 and CAT activity was measured after serum induction. The results are presented as per cent conversion to acetylated form of chloramphenicol. Each sample was independently tested a minimum of three times. *b*, Transcriptional transactivation by Fos proteins. Cotransfection experiments using 5XTRE-TK-CAT as a reporter plasmid were carried out in F9 cells using 4 μ g of reporter plasmid, 5 μ g of c-Jun expression vector, and 6 μ g of each of the *FOS* expression plasmids as indicated. The degree of transactivation compared with the reporter construct plus c-jun is indicated in the lower panel. This experiment was carried out independently at least six times.

METHODS. FosSerA (acidic) mutants were generated by using the phagemid system¹⁶ utilizing the oligodeoxynucleotide shown (right). pSV-c-Jun encodes the murine c-jun cDNA expressed from the SV40 early promoter¹⁹. Transfection as described in Fig. 1 legend.



1,228
5' CACCGCAAGGGC **GAC GAA GAA** AATGAGCCTCC 1,260
Asp Glu Glu

truncated FosSerA seem to be equimolar (compare lanes 7 and 10). As expected, treatment with CIP, which hydrolyses protein phosphoesters, removed all the ^{32}P label (lane 11), which was restored when hydrolysis was carried out in the presence of inhibitor of alkaline phosphatase (lane 12). The combined results of Fig. 2a and b show that FosSerA mutant protein is hypophosphorylated *in vivo* when compared with the wild-type Fos protein.

To understand the mechanism by which the C terminal end of the Fos protein may participate in suppression of *c-fos* transcription, we asked whether the role of phosphorylation was to confer a net negative charge to the C terminus. Consequently, we generated a Fos mutant protein in which the serines at positions 362–364, presumably the sites of Fos phosphorylation, were mutated to aspartic and glutamic acid (FosSerA (acidic)). On transfection of FosSerA (acidic) mutant, the *c-fos* promoter activity was repressed to an extent similar to that observed with the wild-type Fos protein (Fig. 3a: compare lanes 2 and 4). It thus seems that a negative charge at the C terminus of the Fos protein is an important requirement for repressor activity and further strengthens the notion that phosphorylation of the Fos protein may be providing that function.

To confirm that the FosSerA mutant protein is still functional, we asked if it retains its ability to transactivate the transcription of promoters containing a phorbol response element (TRE). Figure 3b shows that in association with Jun, both the wild-type and FosSerA mutant proteins can efficiently transactivate transcription of the *CAT* gene linked to TRE (compare lanes 2 and 3). As previously shown, ν -Fos protein is a potent transcriptional transactivator (lane 4)⁶. Thus, we conclude that: (1) mutant FosSerA protein is functional as judged by its transactivation properties; (2) transrepression and transactivation are two separate activities of the Fos protein; (3) transactivation by Fos protein does not require phosphorylation of the C terminus; and (4) transrepression by Fos protein requires phosphorylation of serine residues in the consensus recognition site of cAMP-dependent protein kinase in the C terminus.

The versatility of the Fos protein as a modulator of transcription is evidenced by its dual function as a transactivator and transrepressor⁶. Mutants in which serine residues have been altered to alanine show little or no transrepression of the *fos* promoter (Fig. 1b) but retain their transactivating properties (Fig. 3b). In contrast, mutation of Ala 356 and 357 to valine had little effect on transrepression (Fig. 1b). We do not know which serine residues in the *c-Fos* C terminus are phosphorylated *in vivo*, but it is clear that the FosSerA mutant is not phosphorylated at the C terminus *in vivo* (Fig. 2). Mutations of individual serine residues (amino acids 362 and 364) in preliminary experiments give proteins that show only partial repression, indicating the need to alter all three serine residues in the *fosA* mutant to abolish transrepression.

We have no firm grip on the precise mechanism of transrepression by Fos, but certain parameters can be established: (1) an intact C terminus is required¹³; (2) C-terminal phosphorylation is required (Fig. 1), but this requirement can be offset by introducing negative charge into the C terminus (Fig. 3); (3) an intact Fos leucine zipper domain is required⁷; (4) the basic region of the Fos protein implicated in binding to TRE is dispensable for transrepression⁷; (5) both the AP-1 site and the cAMP response elements in the *c-fos* promoter are dispensable for repression by Fos protein; and (6) the C terminus alone is not sufficient to cause transrepression because a GAL-4 (1–147 DNA-binding domain) carboxy terminus of *c-Fos* (amino acids 350–380) fusion protein is unable to suppress transcription of the *fos* promoter (R.O. and V.J.D., unpublished results). □

- Chiu, R. *et al. Cell* **54**, 541–552 (1988).
- Sassone-Corsi, P., Lamph, W. W., Kamps, M. & Verma, I. M. *Cell* **54**, 553–560 (1988).
- Lucibello, F. C. *et al. Oncogene* **3**, 43–51 (1988).
- Sassone-Corsi, P., Sisson, J. C. & Verma, I. M. *Nature* **334**, 314–319 (1988).
- Lucibello, F. C., Lowag, C., Neubergh, M. & Muller, R. *Cell* **59**, 999–1007 (1988).
- Van Beveren, C. *et al. Cell* **32**, 1241–1255 (1983).
- Curran, T., Miller, A. D., Zokos, L. & Verma, I. M. *Cell* **36**, 259–268 (1984).
- Barber, J. R. & Verma, I. M. *Molec. cell. Biol.* **7**, 2201–2211 (1987).
- Kemp, E., Graves, D. J., Benjamini, E. & Krebs, E. G. *J. Biol. Chem.* **262**, 4888–4894 (1977).
- De Togni, P. *et al. Molec. cell. Biol.* **8**, 2251–2256 (1988).
- Wilson, T. & Treisman, R. *EMBO J.* **7**, 4193–4202 (1988).
- Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044–1051 (1982).
- Kunkel, T. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 488–492 (1985).
- McClary, J. A., Witney, F. & Geisselsoder, J. *Biotechniques* **7**, 282–284 (1989).
- Deschamps, J., Meijlink, F. & Verma, I. M. *Science* **230**, 1174–1177 (1985).
- Curran, T. & Verma, I. M. *Virology* **135**, 218–228 (1984).
- Lamph, W. W., Wamsley, P., Sassone-Corsi, P. & Verma, I. M. *Nature* **334**, 629–631 (1988).

ACKNOWLEDGEMENTS. We thank P. Sassone-Corsi, J. Barber, I. Tratner, L. Ransone, B. Lamph and T. Hunter for reagents and advice during various stages of this work. We thank P. McClintock for help in preparing the manuscript. R.O. was supported by a long-term EMBO Fellowship. This work was supported by the NIH and the American Cancer Society (I.M.V.).

TATA-dependent and TATA-independent transcription at the *HIS4* gene of yeast

David Pellman*†, Margaret E. McLaughlin* & Gerald R. Fink*‡

* Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

† The Children's Hospital, Boston, Massachusetts 02115, and Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA

Two systems of transcription factors stimulate expression of the *HIS4* gene of *Saccharomyces cerevisiae*. High-level transcription induced by amino-acid starvation is mediated by Gcn4 (ref. 1) and basal transcription is mediated jointly by the factors Bas1 and Bas2 (Pho2)^{2–4}. We now show that wild-type Gcn4 requires the TATA element for correct messenger RNA start-site selection, but Gcn4 derivatives with deletions in the activation domain activate *HIS4* transcription at the correct mRNA start site (I) in the absence of the TATA element. Gcn4 derivatives that activate TATA-independent transcription show low levels of activation. Similarly, we find that low levels of transcription by Bas1/Bas2 are TATA-independent, whereas high levels are TATA-dependent. These results show that the *HIS4* TATA element is required for mRNA start-site selection only under conditions of high-level transcription.

The *HIS4* promoter has a single functional TATA element located at position –123 from the A of the ATG of the initiator methionine⁵. The *HIS4* promoter also contains binding sites for the transcriptional activators Gcn4, Bas1 and Bas2^{3,5,6}, and the mRNA initiation site at position –63 (ref. 5) (Fig. 1a). In the studies that follow, the TATA requirement of various *HIS4* activators was assayed from two different TATA mutants. The first mutant, *his4TATAΔ*, is a deletion-insertion encompassing the –123 TATA element^{5,7}. The second mutant, *his4TATA–*, has point mutations in the –123 TATA as well as in the three other TATA-containing sequences (Fig. 1a).

A simple Petri plate assay reveals the effect of the TATA element on *HIS4* expression stimulated by various transcription factors. A strain lacking all three *trans*-acting factors (*bas1 bas2 gcn4*) is His[–] (fails to grow without histidine; Fig. 1b, control), as a consequence of diminished *HIS4* transcription from the normal mRNA start at –63 (ref. 7). In the presence of a functional TATA –123, strains containing either Gcn4 (*bas1 bas2 GCN4*) or Bas1/Bas2 (*BAS1 BAS2 gcn4*) are His[–], showing that either system can stimulate *HIS4* expression. In the absence

of a functional TATA, a *bas1 bas2 GCN4* strain is His⁻, whereas a *BAS1 BAS2 gcn4* strain is His⁺. This comparison shows that Gcn4 has a strong requirement for the TATA-123, whereas Bas1/Bas2-activated transcription is TATA-independent (Fig. 1b).

Analysis of the *HIS4* mRNA start sites shows that Gcn4 requires the -123 TATA element for correct initiation from the -63 site. Both *HIS4* TATA element mutations cause an overall reduction in transcription stimulated by Gcn4 and an alteration in start-site selection: there is a roughly 20-fold reduction in transcripts initiated at the -63 site and the majority of the remaining transcripts initiate at several downstream sites within the *HIS4* coding sequence (Fig. 2, lanes 4-6). By contrast, TATA element mutations reduce the amount of transcription induced by Bas1/Bas2 only moderately and do not alter the site of mRNA initiation (Fig. 2, lanes 7-9).

The different requirements of Bas1/Bas2 and Gcn4 for a TATA element could reflect either different structural features of the factors, or a difference between low-level and high-level transcription. Bas1/Bas2-activated transcription can be increased by growing cells in adenine-deficient medium³ (Fig. 2; compare lanes 7, 10, 13). In *HIS4* TATA mutation strains grown under these conditions, transcripts activated by Bas1/Bas2 initiate at multiple sites in a pattern similar to that shown by transcripts activated by Gcn4 from TATA-less promoters (Fig. 2, lanes 14, 15). Therefore, low levels of transcription activated by Bas1/Bas2 are TATA-independent, but high-level transcription requires a TATA element for accurate initiation.

If the distinction between TATA-dependent and TATA-independent transcription at *HIS4* is related to the strength of activation, low-level activation by Gcn4 should not require a TATA element. Gcn4 is a 281-amino-acid protein consisting of two domains: a DNA-binding domain located in the 60 amino acids at the C terminus⁸, and an acidic activation domain. The number of acidic amino acids in the activation domain—

previously mapped to ~50 amino acids in the middle of the protein^{9,10}—was shown to be roughly proportional to the strength of activation. To test the effect of Gcn4 activation strength on TATA dependence, we compared the ability of a series of Gcn4 proteins with deletions in the activation domain to stimulate *HIS4* transcription in the presence and absence of the TATA-123. The *GCN4* constructs were first tested by Petri plate assay. In strains containing a functional TATA (Fig. 3), a Gcn4 deletion of up to 95 amino acids (*gcn4*-C186) has no apparent effect on activation strength, deletion of 117 amino acids produces a weaker activator (*gcn4*-C163), and deletion of 140 amino acids removes the activation function (*gcn4*-C141, not shown)⁹. When this same deletion series was tested in a strain containing the *his4TATAΔ* allele, a strikingly different result was obtained. Wild-type (wt) Gcn4 or deletions of up to 84 amino acids (*gcn4*-C197) are unable to activate the *his4TATAΔ* allele. But deletions that remove a greater number of acidic amino acids, *gcn4*-C186, *gcn4*-C163 and *gcn4*-Δ34, activate without the TATA element to give a His⁺ phenotype (Fig. 3).

The primer extension analysis shows that weak Gcn4 activators can stimulate TATA-independent transcription. The construct *gcn4*-C163, which activates at about the same level as basal transcription by Bas1/Bas2, is independent of the *HIS4* TATA element, both with respect to the overall level of transcription and with respect to the site of mRNA initiation (Fig. 4, lanes 10-12). The constructs *gcn4*-C186, *gcn4*-C243 and *gcn4*-Δ34 are intermediate in strength and also intermediate for the requirement for a TATA element; these deletion proteins stimulate a higher ratio of correct to incorrect initiation without a TATA element than Gcn4-wt (Fig. 4). The fact that deletions in non-overlapping regions of the activation domain have the same effect on the TATA requirement (compare *gcn4*-C186 with *gcn4*-Δ34) supports the idea that the TATA-independence of the Gcn4 deletions is a direct consequence of diminished activation.

FIG. 1 a, The *HIS4* promoter. *HIS4* mRNA initiates at position -63 from the A of the codon for the initiator methionine. The open arrows indicate the Gcn4 consensus binding sites; the site of the Gcn4 icon is the main site^{5,6}. The size of the Gcn4, Bas1 and Bas2 icons approximate the size and location of the corresponding DNase I footprint^{3,6}. The *HIS4* TATA at -123 (TATATAA) is similar in sequence and position to other yeast TATA elements²⁰; its deletion dramatically reduces transcription. The *HIS4* promoter contains the shorter DNA sequence (TATA) at three other positions, but deletion of these sequences has no effect on *HIS4* expression⁵. The DNA sequence shown is from nucleotides -96 to -133 and includes three of the four TATAs. The *his4TATA* allele was constructed by site-directed mutagenesis of a 1.1 kilobase (kb) *PvuII*-*SalI* fragment of *HIS4*, which was then integrated to replace the wt *HIS4* promoter. The *his4TATA* allele contains the following base changes: -167, A→C; -130, T→G; -121, T→G; -119, T→G; -100, T→G; -89, G→C; -85, T→G, creating *EcoRV* and *NcoI* recognition sites at -172 and -89. b, Histidine phenotypes with or without the -123 TATA element. Strains containing only Bas1 or Bas2 are His⁻ because both Bas1 and Bas2 are required for any basal transcription^{2,3}. Growth is shown on an SC-his-ura plate⁷. The control strain, L3110 (*MATα bas1 bas2 gcn4 HIS4 ura3*-52) with plasmid YCp88 (ref. 9) is His⁺. The *bas1 bas2 gcn4 his4TATAΔ* strains are also His⁻. The *bas1*, *bas2* and *gcn4* alleles are all deletion alleles^{2,7}. The *GCN4* strains are: L3110 (YCp88 - *GCN4*) and L3412, (*MATα bas1 bas2 gcn4 his4TATAΔ ura3*-52 *ino1*-13 (YCp88 - *GCN4*)). The *BAS1/BAS2* strains are: L3079, (*MATα gcn4 ura3*-52 *ino1*-13 (YCp88)) and L4389 (*MATα gcn4 his4 TATAΔ ura3*-52 *ino1*-13 (YCp88)).

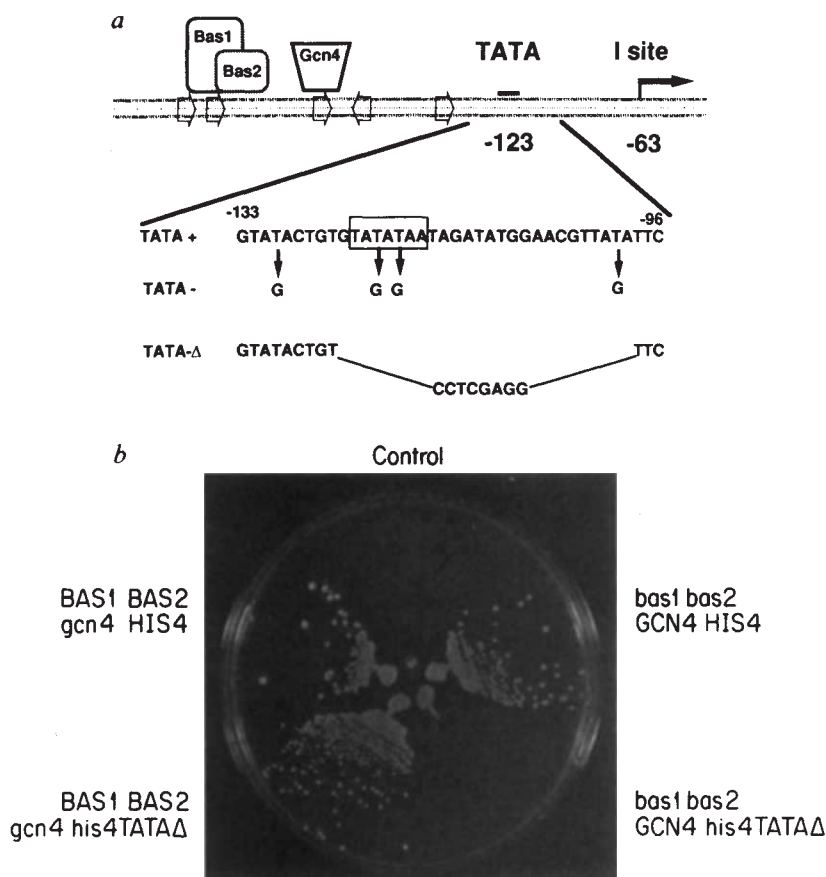
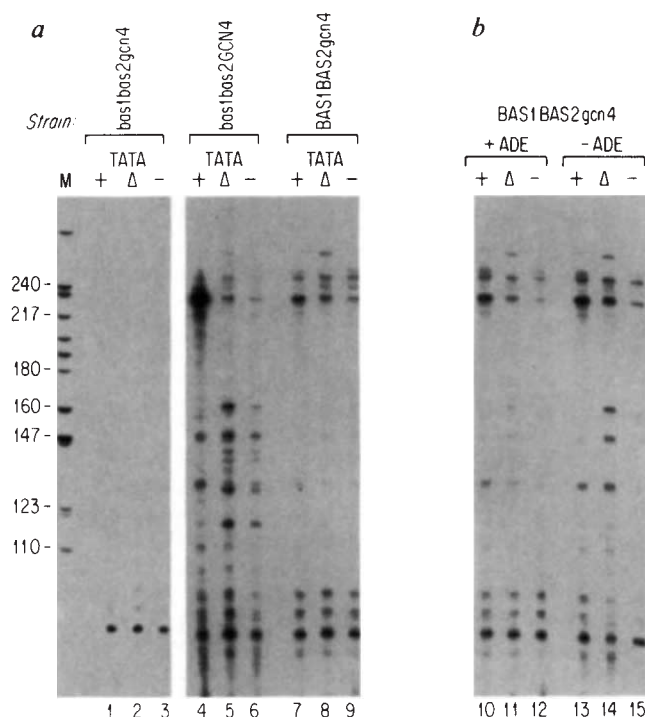


FIG. 2 A comparison of *HIS4* transcripts stimulated by Gcn4-wt and Bas1/Bas2. Each lane shows primer extensions from 25 µg total RNA annealed to a 70-base *HIS4* primer and a 21-base *CYH2* primer (to control for the amount of RNA)²¹. The lanes are grouped into sets of three: TATA+, TATAΔ and TATA-. a, Comparison of expression stimulated by Gcn4 and Bas1/Bas2. RNA was prepared from cells grown in SC-ura (ref. 7). The strains from lanes 1–9, are: 1, L3110(YCp88); 2, L3412 (YCp88); 3, L4404 (*MATα bas1 bas2 gcn4 his4 TATA_ura3-52 leu2-112*) (YCp88); 4, L3110 (YCp88-GCN4); 5, L3412(YCp88-GCN4); 6, L4404(YCp88-GCN4); 7, L3079(YCp88); 8, L4389(YCp88) and 9, 10166-B (*MATα gcn4 his4TATA-ura3-52 leu2-112*) (YCp88). For genotypes of L3110, L3412, L3079 and L4389 see Fig. 1. The level of transcription correlates with the growth phenotype: *BAS1 BAS2 gcn4* with (wt) TATA-123 has a stronger His⁺ phenotype than *BAS1 BAS2 gcn4* with the TATAΔ (Fig. 1b). A *BAS1 BAS2 gcn4 his4TATAΔ* strain synthesizes only slightly more correctly initiated transcripts than the *bas1 bas2 GCN4 his4TATAΔ* strain, yet the former grows without histidine whereas the latter does not. This difference in transcription from the -63 site therefore defines a threshold level of message below which the phenotype is His⁻ and above which the phenotype is His⁺. b, Effect of TATA mutations on high level transcription by Bas1/Bas2. Lanes 10–12 show primer extension products from RNA from L3079, L4389 and 10166-B grown in minimal +ade (SD + ura + his + leu + arg + ino + ade). Lanes 13–15 show the primer extension products from RNA from the same strains grown without adenine. All lanes from this figure are from the same exposure of the same gel.

METHODS. Total RNA was isolated from strains grown to Klett 40 (ref. 22). The 70-base *HIS4* primer (complementary to sequences +99 to +169) was prepared by Klenow enzyme extension of a 22-base oligonucleotide (+147 to +169) hybridized to single-stranded M13 DNA containing *HIS4* sequences -588 to +469. Extension was done in the presence of [α -³²P]dGTP, the product was digested with *Xho*I, and then isolated from a 6% polyacrylamide denaturing gel. The 21-base *CYH2* primer, end-labelled with ³²P, is complementary to sequences +564 to +587 of *CYH2* (ref. 21). For the primer extension reactions the *HIS4* primer (500,000 c.p.m.) and the *CYH2* primer (50,000 c.p.m., diluted 1 to 10 with cold primer) were mixed with 25 µg of total RNA, hybridized, and extended with reverse transcriptase²³. The reac-



tion products were separated on a 6% polyacrylamide denaturing gel. The marker (M) is an *Msp*I digest of pBR322, labelled with [α -³²P]dCTP (sizes in kb). The positions of the four major internal initiation sites in lane 5 (+10, +22, +46 and +53) were determined by running a similar sample next to a DNA sequencing ladder.

TATA

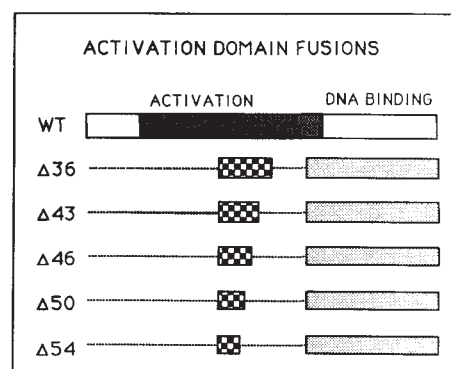
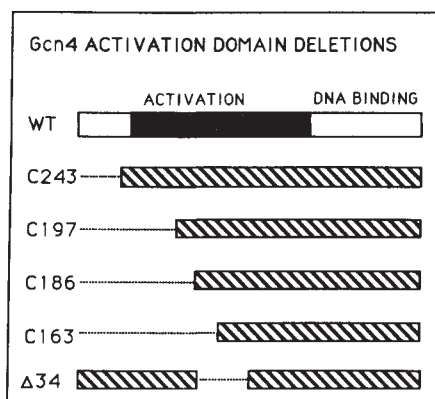
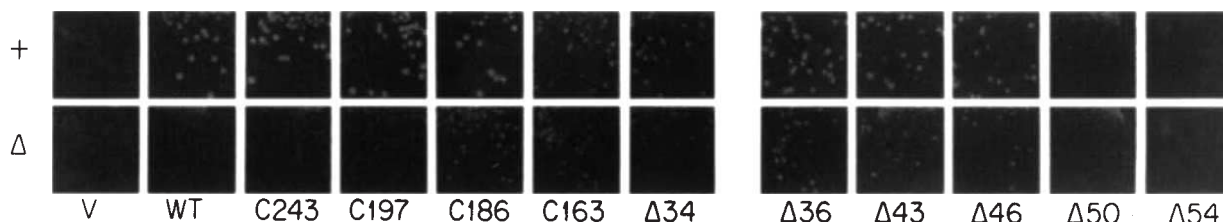
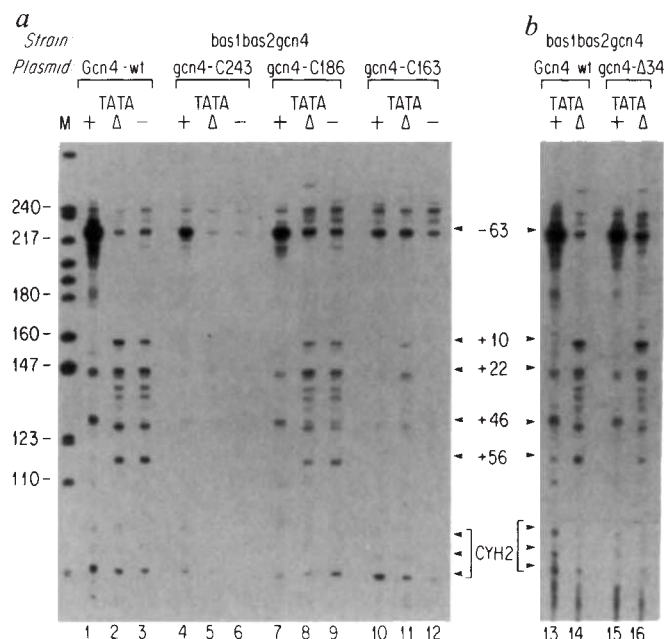


FIG. 3 TATA dependence of *HIS4* expression activated by Gcn4. All *GCN4* constructs were cloned into YCp88 and expressed constitutively from the *DED1* promoter^{9,10}. YCp88-gcn4Δ34 was constructed from YCp88-GCN4 and YCp88-LexA-gcn4Δ34 (K. Struhl, unpublished data). The dotted lines indicate the deletions and the bars represent the amino acids remaining in each Gcn4 construct. The shaded area of the map of Gcn4-wt indicates the

area rich in acidic amino acids. The strains are: L3110 and L3412. The colonies shown have been grown for 2.5 days at 30 °C SC-his-ura. Both strains with the different GCN4 plasmids have similar rates on SC-ura. Growth of L4404 on SC-ura-his (which contains the *his4TATA* - allele) with the various *GCN4* constructs parallels that of L3412 (*his4TATAΔ*). But in L4404, Gcn4-wt does not induce as strong a His⁻ phenotype as in L3412.

FIG. 4 The effect of TATA mutations on *HIS4* expression induced by Gcn4 deletions. *a*, N-terminal deletions of Gcn4. The lanes are grouped into sets of three according to strain: L3110 (TATA+), L3412 (TATA Δ) and L4404 (TATA-). Each strain has been transformed with the indicated plasmid. The plasmids are described in the legend to Fig. 3 and the genotypes of the strains are given in the legends to Figs 1 and 2. The strains were grown in SC-ura and primer extension analysis was performed as described in Fig. 2. Transcription from the wt *HIS4* allele provides an estimate of the activity of each Gcn4 mutant; gcn4-C243 is a weaker activator than some of the constructs with larger deletions, possibly because it is less stable. *b*, An internal deletion of Gcn4. The strains are L3110 (lanes 13 and 14) and L3412 (lanes 15 and 16). The *CYH2* transcripts for lanes 13–16 were visualized by a longer exposure of this portion of the gel (4 days rather than 1). Size markers (M) on left-hand side (in kb).



Analysis of another series of *GCN4* constructs (activation domain fusions; Fig. 3)¹⁰ shows that the strength of TATA-independent activation is proportional to the size of the activation domain. The proteins encoded by these constructs have small segments of the activation domain fused to the DNA-binding domain. All of these constructs activate the *his4TATA Δ* allele, and the strength of activation is proportional to the size of the N-terminal acidic sequence. The parallel in activation strength from the wild-type *HIS4* promoter and the TATA deletion promoter suggests that the activation domain functions similarly with or without the TATA element.

Our results shed new light on the extent of the Gcn4 activation domain. Quantitative assay of activation strength by primer extension analysis shows that every deletion in the N-terminal half of the protein produces a weaker activator than Gcn4-wt (Fig. 4, lanes 1, 4, 7, 10 and 15). We therefore conclude that the amino terminus of Gcn4 contains sequences that should be considered part of the activation domain. The analysis of *gcn4- Δ 34* supports this conclusion. The mutant *gcn4- Δ 34* lacks the region of Gcn4 previously suggested to be of primary importance for activation¹⁰, yet activates at about half the level of Gcn4-wt. The simplest explanation for the ability of *gcn4- Δ 34* to activate is that the amino-terminal portion of Gcn4 provides the activation function in the absence of the internal activation domain. The fact that the N-terminal portion of Gcn4 contributes to activation is less surprising, given that the entire region (from amino acid numbers 50 to 180) is rich in acidic amino acids.

Our finding that the discrimination between TATA-dependent and TATA-independent transcription at *HIS4* is due to the strength of activation contrasts with results from other studies, which suggest that it is the transcription factors themselves that are 'TATA-dependent' or 'TATA-independent'^{11–13}. For both Gcn4 and Bas1/Bas2, *HIS4* transcription is TATA-independent at low levels and TATA-dependent at high levels. The correlation between activator strength and the ability to stimulate TATA-independent transcription is illustrated most convincingly by the Gcn4 activation domain deletions. As activation strength is reduced, the proportion of correctly initiated transcripts from the *HIS4* TATA mutations increases. These results suggest that there is a fundamental difference between high and low level transcription; high-level transcription requires the TATA element whereas low-level transcription does not. We note that transcription from many mammalian TATA-less promoters is at low levels^{14–16}.

How does accurate initiation occur without a TATA element? One possibility is that for basal transcription, the site of mRNA

initiation (I site) directs the assembly of activators with general transcription factors. The *HIS4* I site is required for accurate start-site selection. Deletion of the I site or alteration of the spacing between the I site and the TATA element results in heterogeneous transcription initiation⁵. Furthermore, the I site is critical for mRNA initiation in a number of yeast promoters^{5,17,18,24}. We now find that accurate use of the *HIS4* I site can occur without a TATA element. The *HIS4* I site may therefore be analogous to the initiator from the mammalian terminal deoxynucleotidyltransferase gene, which can function with or without a TATA element^{14,19}. Perhaps at low levels of *HIS4* expression, activators can assemble with factors that are bound at the I site, but at high levels of transcription, efficient assembly at the I site additionally requires a TATA element. In the absence of the TATA element, strong activators like Gcn4-wt may scan downstream indiscriminately recognizing weak TATA elements, resulting in initiation within the *HIS4* coding sequence. This model implies that the I site may be the primary determinant of mRNA start site selection at *HIS4*, with the TATA element required only to direct strong activation from the I site. □

Received 25 June; accepted 20 August 1990.

- Hinnebusch, A. G. *Microbiol. Rev.* **52**, 248–273 (1988).
- Arndt, K. T., Styles, C. & Fink, G. R. *Science* **237**, 874–880 (1987).
- Tice, B. K., Fink, G. R. & Arndt, K. T. *Science* **246**, 931–935 (1989).
- Sengstag, C. & Hinnebusch, A. *Nucleic Acids Res.* **15**, 233–246 (1987).
- Nagawa, F. & Fink, G. R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8557–8561 (1985).
- Arndt, K. & Fink, G. R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8516–8520 (1986).
- Arndt, K. T., Styles, C. A. & Fink, G. R. *Cell* **56**, 527–537 (1989).
- Hope, I. A. & Struhl, K. *Cell* **43**, 177–188 (1985).
- Hope, I. A. & Struhl, K. *Cell* **46**, 885–894 (1986).
- Hope, I. A., Mahadevan, S. & Struhl, K. *Nature* **333**, 635–640 (1988).
- Greene, J. M. & Kingston, R. E. *Molec. cell. Biol.* **10**, 1319–1328 (1990).
- Taylor, I. C. & Kingston, R. E. *Molec. cell. Biol.* **10**, 165–175 (1990).
- Wefald, F. C., Devlin, B. H. & Williams, R. S. *Nature* **344**, 260–262 (1990).
- Smale, S. T. & Baltimore, D. *Cell* **57**, 103–113 (1989).
- Reynolds, G. et al. *Cell* **38**, 275–286 (1984).
- Sehgal, A., Patil, N. & Chao, M. *Molec. cell. Biol.* **8**, 3160–3167 (1988).
- Chen, W. & Struhl, K. *EMBO J.* **4**, 3273–3280 (1985).
- McNeil, J. B. & Smith, M. *Molec. cell. Biol.* **5**, 3545–3551 (1985).
- Smale, S. T., Schmidt, M. C., Berk, A. J. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4509–4513 (1990).
- Struhl, K. A. *Rev. Biochem.* **58**, 1051–1077 (1989).
- Kaufman, N. F., Fried, H. M., Schwindinger, W. F., Jasini, M. & Warner, J. R. *Nucleic Acids Res.* **11**, 3123–3135 (1983).
- Elion, E. A. & Warner, J. R. *Cell* **39**, 663–673 (1984).
- Lillie, J. W., Green, M. & Green, M. R. *Cell* **46**, 1043–1051 (1986).
- Hahn, S., Hoar, E. & Guarente, L. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8562–8566 (1985).

ACKNOWLEDGEMENTS. We thank K. Struhl for the gift of the *GCN4* deletion plasmids; D. Hekmatpanah for strains and plasmids; C. Styles and P. Grisafi for technical assistance; K. Struhl, S. Smale and members of our laboratory for discussion; and M. Christman, A. D'Andrea, E. Elion, D. Miller, H. Rudolph and B. Ruskin for comments on the manuscript. D.P. is a St. Jude's Children's Hospital fellow in the Pediatric Scientist Training Program. This work was supported by NIH grants to G.R.F. G.R.F. is American Cancer Society Professor of Genetics.

NATURE • VOL 348 • 1 NOVEMBER 1990

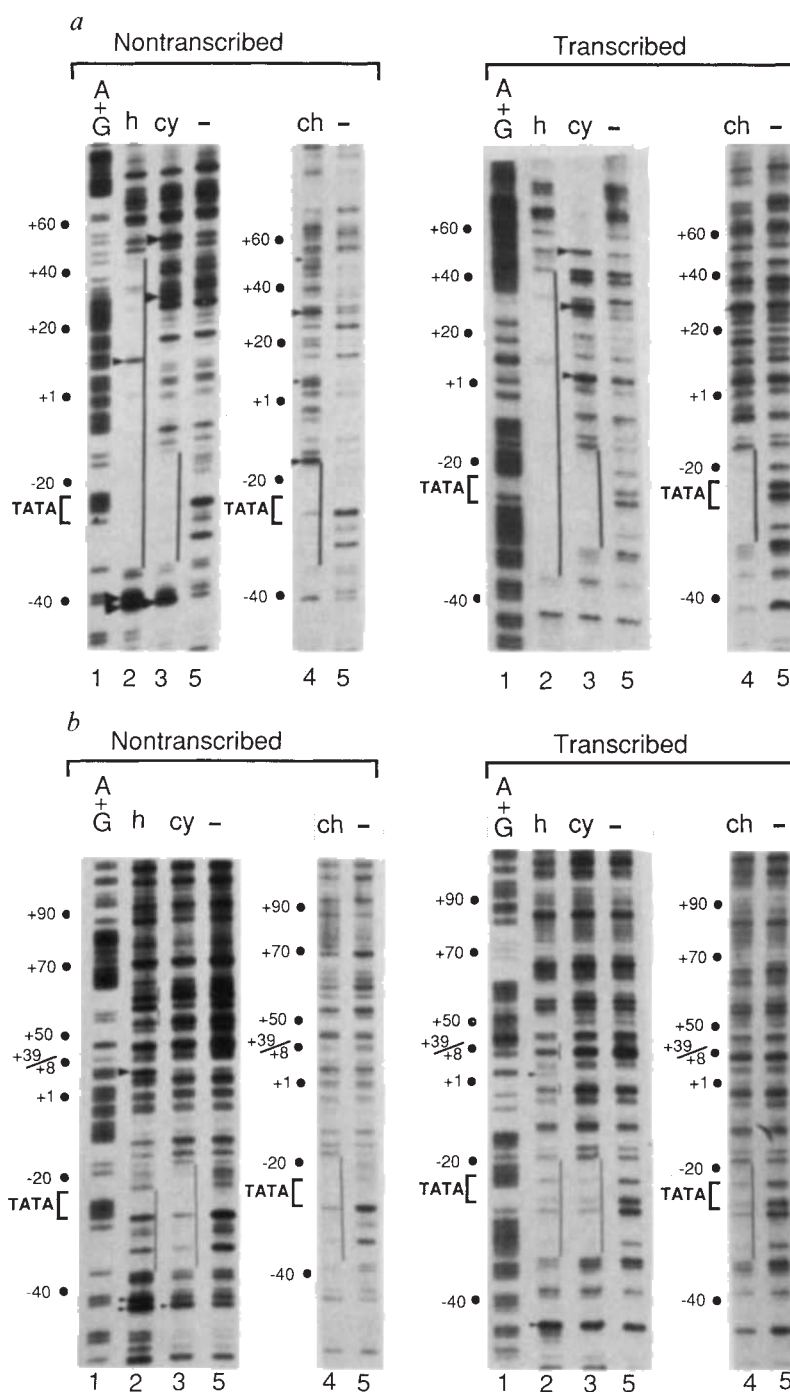


FIG. 2 DNase I footprinting analysis of TFIID binding to the *gfa* promoter. The WT (a) and Δ DN (b) probes were used for DNase I footprinting either with hTFIID (lanes 2), cyTFIID (lanes 3) or chTFIID (lanes 4) or without TFIID (lanes 5). The A+G-specific chemical cleavage reactions were used as molecular mass markers (lanes 1). Arrows point to the hypersensitive sites.

METHODS. Plasmid pGEM-3Z (Promega) containing a *gfa* *SacI* fragment (positions -132 to +242) having either WT or Δ DN sequences (Fig. 1) was digested with *EcoRI* (adjacent to the upstream *SacI* site at -132) and either the non-transcribed or transcribed strand was labelled using polynucleotide kinase or the Klenow fragment of DNA polymerase I, respectively. After digestion with *HindIII* (adjacent to downstream *SacI* site at +242), the smaller *gfa*-containing fragment was purified and used as a probe for footprinting. The specific activities of the probes were about 10,000 d.p.m. fmol⁻¹. DNase I footprinting was carried out as described¹¹.

probe, both the chTFIID footprint and the hypersensitive site at position -15 were slightly weaker than those observed with the WT probe (Fig. 2a and b, lanes 4). Moreover, the downstream hypersensitive sites at positions +6, +28 and +50 were no longer apparent. The results with cyTFIID were essentially identical to those with chTFIID except that the downstream hypersensitive sites in the WT probe appeared at positions +3, +28 and +46 (Fig. 2a and b, lanes 3). Although these results suggest that the downstream sequence may contribute to stable chTFIID or cyTFIID binding, the overall effect of the downstream sequence on TFIID binding was much weaker for chTFIID or cyTFIID than for hTFIID.

We exploited the difference between hTFIID and chTFIID or cyTFIID binding to ask whether the effect of the downstream initiator on *gfa* gene transcription was due to its contribution to TFIID binding. If this were so, then deletion of the downstream initiator should inhibit transcription mediated by hTFIID

more strongly than that mediated by chTFIID or cyTFIID. To test this hypothesis, we used a nuclear extract that had been depleted of endogenous TFIID activity by heat treatment³. The WT and Δ DN templates were cotranscribed in the presence or absence of varying amounts of hTFIID, chTFIID and cyTFIID, and the transcripts detected by primer extension analysis (Fig. 3). When the extract was supplemented with hTFIID, transcription from the Δ DN template was reduced 30-fold relative to that from the WT template. In marked contrast, when chTFIID and cyTFIID were used, transcription from the Δ DN template was reduced only threefold relative to the WT template. This 10-fold difference in the relative activities was not a function of the particular level of each TFIID, but rather was due to an intrinsic difference between hTFIID and the cloned factors (chTFIID and cyTFIID). Together with the footprinting data (Fig. 2), these results indicate that the downstream initiator stimulates transcription by contributing to TFIID binding, most

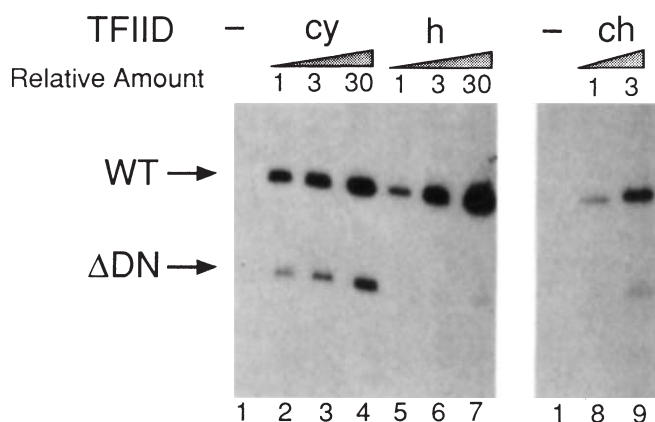


FIG. 3 *In vitro* transcription of *gfa* initiator mutants. The WT and Δ DN templates were cotranscribed in heat-treated nuclear extracts supplemented with buffer (lane 1); cyTFIID (lanes 2–4); hTFIID (lanes 5–7) or chTFIID (lanes 8–9). The expected positions of the reverse transcripts are shown on the left side of the autoradiogram.

METHODS. *In vitro* transcription³ and primer extension¹ were performed as described. Heat-treated nuclear extracts³ were prepared as before.

notably in the case of a natural (partially purified) human factor. Preliminary experiments have shown that the Δ TATA template is transcribed weakly in a TFIID-dependent nuclear extract supplemented with hTFIID, but not with cloned TFIIDs (our unpublished data). Although this result suggests that hTFIID can interact with the downstream initiator in the absence of the TATA box, no footprint was detected on the Δ TATA probe with any of the three TFIID preparations under any of our conditions (data not shown). The hTFIID-downstream initiator interaction

may be weak and/or unstable in the absence of the TATA box or the other general transcription factors.

The different behaviours of natural and cloned factors suggest the presence of an additional component(s) in the hTFIID fraction that copurifies with human TFIID. Such a factor might be a subunit of natural TFIID because natural hTFIID has a higher relative molecular mass of 120,000 (120K)⁶ than chTFIID (37 K)⁴; and because a more highly purified single-stranded DNA agarose hTFIID fraction³ (fifth chromatographic step) gives the same footprinting pattern as the ω -amino octyl agarose hTFIID fraction⁷ (third chromatographic step) used in this study. A model summarizing the binding and transcriptional properties of the natural and cloned TFIIDs on the WT and Δ DN templates is presented in Fig. 4.

A previous study indicated that deletion of either the TATA box or the downstream initiator reduces *gfa* transcription 10-fold *in vitro*¹. Here we have shown that deletion of the downstream initiator also inhibits transcription *in vivo*. The further demonstration that the downstream initiator contributes to TFIID binding to the TATA box provides a molecular explanation for this loss of function. Although several types of initiator elements have now been identified^{8,9}, this is the first demonstration of such an element having an effect on promoter interactions of a known transcription factor. Finally, we note that the additional component(s) associated with TFIID might be important not only for basal level transcription, but also for regulation by *trans*-acting factors. Horikoshi *et al.* demonstrated that some activators (GAL4¹⁰, ATF¹¹) affect the downstream interaction of hTFIID in the adenovirus E4 promoter. In addition, whereas hTFIID mediates activation by upstream activators in a system reconstituted with partially purified general factors, chTFIID supports only basal transcription in the same system⁴. □

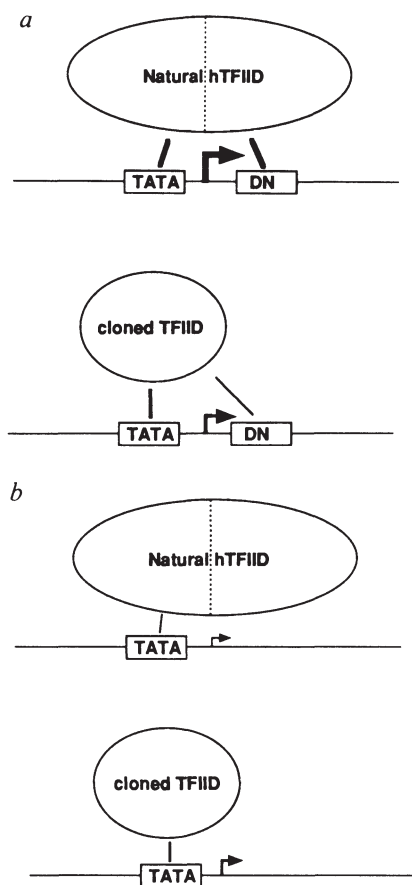


FIG. 4 A possible mechanism for interactions of TFIID with initiation elements. Natural hTFIID binds strongly to the WT promoter (a) through interaction at both the TATA element and the downstream initiator (DN). The relatively weaker binding of natural TFIID to the Δ DN template (b) indicates that interactions at the downstream initiator influence binding at the TATA element. The cloned TFIID component interacts primarily at the TATA element and lacks a component(s) that seems to facilitate TATA element interactions in the presence of the downstream initiator. The thickness of the lines connecting TFIID to the DNA indicates the strength of the interaction.

Received 8 May; accepted 30 August 1990.

1. Nakatani, Y., Brenner, M. & Freese, E. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4289–4293 (1990).
2. Sawadogo, M. & Roeder, R. G. *Cell* **43**, 165–175 (1985).
3. Nakajima, N., Horikoshi, M. & Roeder, R. G. *Molec. cell. Biol.* **8**, 4028–4040 (1988).
4. Hoffmann, A. *et al. Nature* **346**, 387–390 (1990).
5. Horikoshi, M. *et al. Nature* **341**, 299–303 (1989).
6. Reinberg, D., Horikoshi, M. & Roeder, R. G. *J. biol. Chem.* **262**, 3322–3330 (1987).
7. Horikoshi, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 4843–4847 (1989).
8. Smale, S. T. & Baltimore, D. *Cell* **57**, 103–113 (1989).
9. Means, A. L. & Farnham, P. J. *Molec. cell Biol.* **10**, 653–661 (1990).
10. Horikoshi, M., Carey, M. F., Kakidani, H. & Roeder, R. G. *Cell* **54**, 665–669 (1988).
11. Horikoshi, M., Hai, T., Lim, Y. S., Green, M. R. & Roeder, R. G. *Cell* **54**, 1033–1042 (1988).
12. Brenner, M. *et al. Molec. Brain Res.* **7**, 277–286 (1990).
13. Pfrendrich, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 5122–5126 (1978).
14. Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044–1051 (1982).
15. Wigler, M., Pellicer, A., Silverstein, S. & Axel, R. *Cell* **14**, 725–731 (1978).
16. Nordeen, S. K. *Biotechniques* **6**, 454–457 (1988).
17. de Wet, J. R., Wood, K. V., DeLuca, M., Helinski, D. R. & Subramani, S. *Molec. cell. Biol.* **7**, 725–737 (1987).

ACKNOWLEDGEMENTS. We thank D. Schoenberg for editing the manuscript. M.H. is an Alexandrine and Alexander L. Sinsheimer Scholar. A part of this study was supported by the NIH (R.G.R.) and by the Pew Trusts to The Rockefeller University.

Ferryl and hydroxy intermediates in the reaction of oxygen with reduced cytochrome *c* oxidase

Sanghwa Han, Yuan-chin Ching & Denis L. Rousseau

AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA

CYTOCHROME *c* oxidase catalyses the 4-electron reduction of dioxygen to water and translocates protons vectorially across the inner mitochondrial membrane. Proposed reaction pathways for the catalytic cycle of the O₂ reduction¹⁻³ are difficult to verify without knowing the structures of the intermediates, but we now have such information for the catalytic intermediates in the first steps of the reaction of O₂ with cytochrome *c* oxidase from resonance Raman spectroscopy⁴⁻⁶, a technique that enables iron-ligand stretching modes to be identified⁴⁻⁷. Here we report on two more key intermediates: a ferryl-oxo (Fe⁴⁺=O²⁻) and a ferric-hydroxy (Fe³⁺-OH⁻) intermediate at the level of 3- and 4-electron reduction, respectively. We identified these intermediates by their characteristic iron-oxygen stretching frequencies (786 cm⁻¹ for Fe⁴⁺=O²⁻, and 450 cm⁻¹ for Fe³⁺-OH⁻) and oxygen and deuterium isotope shifts. The oxo atom in the ferryl intermediate is hydrogen-bonded and the iron-oxygen bond in the hydroxy intermediate is anomalously weak. With the identification of the primary, ferryl and hydroxy intermediates, the predominant structures at almost all stages of O₂ reduction are now known and the catalytic pathway can be described with more certainty.

To identify the oxygen isotope-sensitive lines in the reaction of cytochrome *c* oxidase with ¹⁶O₂ and ¹⁸O₂, the resonance Raman spectrum of the intermediates in the reaction of the enzyme with each isotope was followed from 15 μs to 3 ms using a flow-flash-probe method. In this method, a continuously flowing sample of carbon monoxide-bound cytochrome *c* oxidase is photodissociated by one laser beam and the reaction products are probed by another⁸. Three oxygen isotope-sensitive lines were detected (Fig. 1; Table 1). At the earliest times in the reaction of fully reduced cytochrome *c* oxidase with ¹⁶O₂, a line at 568 cm⁻¹, with a 21-cm⁻¹ oxygen isotope shift, has been assigned as the Fe-O₂ stretching mode of the primary intermediate^{4-6,8}. After this mode decays a line is detected at 786 cm⁻¹, and at the longest times a line appears at 450 cm⁻¹ (Figs 1 and 2). No isotope-sensitive lines, other than that at 568 cm⁻¹ were detected in similar experiments using the mixed-valence form of the enzyme.

The second oxygen isotope-sensitive line at 786 cm⁻¹ has a 35-cm⁻¹ oxygen isotope shift. It is present in the spectra from 300 to 1,500 μs (Fig. 2) and reaches its maximum intensity within 500 μs. On the basis of reported kinetic constants^{1,9} and Raman measurements of the oxidation of cytochrome *a* (ref. 10), the intermediate is assigned to the 3-electron reduction level (one electron from cytochrome *a*₃, one from Cu_B, which along with cytochrome *a*₃ forms the binuclear O₂ binding site, and one from the low-potential centres, cytochrome *a* and Cu_A). To test whether the molecular entity with the 786-cm⁻¹ mode is bound to or interacts with protons, the reaction with oxygen was also carried out in buffered D₂O (Fig. 1). Because the reaction rate in D₂O is different from that in H₂O, the effect of D₂O was

determined by comparing the reaction of cytochrome *c* oxidase with ¹⁶O₂ and with ¹⁸O₂ in the deuterated buffer. As shown in Table 1, in D₂O this mode shifts to higher frequency by 12–15 cm⁻¹ for both oxygen isotopes. In addition, the intensity of this isotope-sensitive line is stronger in D₂O than in H₂O and persists for longer.

We identify the line at 786 cm⁻¹ as an iron-oxygen stretching mode of a ferryl (Fe⁴⁺=O²⁻) intermediate because of the similarity of its frequency and oxygen isotope shift (theoretical shift, 35 cm⁻¹) to those of other reported ferryl complexes (Table 2). An experiment with singly substituted molecular oxygen (¹⁶O-¹⁸O) confirms that the line at 786 cm⁻¹ is a ferryl (single oxygen) and not a peroxide (double oxygen) intermediate (data not shown). If the normal coordinate of a vibrational mode includes proton contributions, the increase in mass due to deuterium substitution is expected to lead to a decrease in frequency. But we detect a clear and large increase in frequency upon deuteration in the reaction with both ¹⁶O₂ and ¹⁸O₂ for this ferryl mode, demonstrating that an exchangeable proton is not bound to the oxygen atom.

The effect of pH and exposure to deuterated buffers on ferryl-oxo stretching modes has been studied¹¹⁻¹⁷ in other haem proteins (Table 2). In several cases, a large increase in frequency (10–15 cm⁻¹) is observed when the pH becomes alkaline. This has been interpreted as resulting from a loss in hydrogen bonding to the ferryl-oxo atom due to deprotonation of an amino-acid

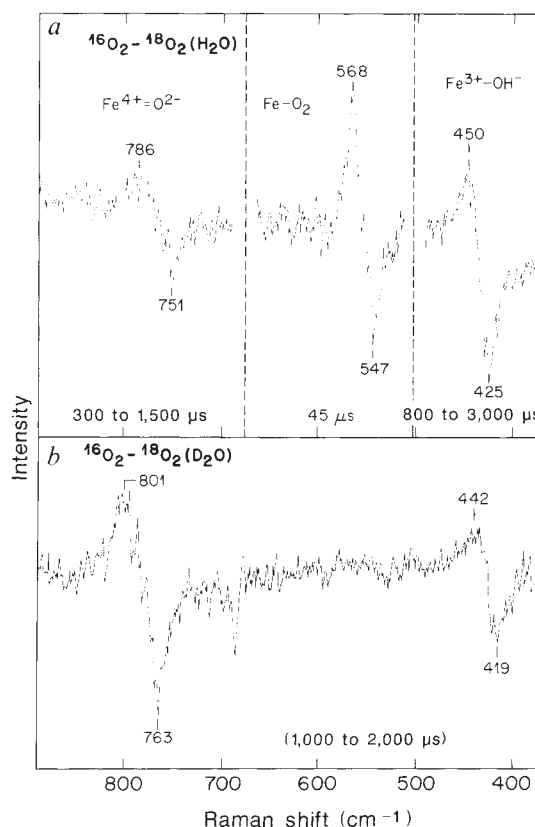


FIG. 1 Isotope difference spectra (¹⁶O₂-¹⁸O₂) at selected time ranges after the reaction of oxygen with cytochrome *c* oxidase. Several spectra obtained throughout the indicated time ranges were added together to improve the signal-to-noise ratio. *a*, Measurements in buffered H₂O showing the three oxygen isotope-sensitive lines detected over the 45 to 3,000 μs time range. *b*, Measurements in buffer of D₂O in a time range in which the ferryl (801 cm⁻¹) and the hydroxy (442 cm⁻¹) intermediates are strong. The enzyme (200 μM before mixing), isolated by the method of Yonetani²⁸, was solubilized in phosphate buffer (100 mM, pH 7.4) with 1% dodecyl β-D-maltoside. It was reduced with 20 mM ascorbate and catalytic amounts of cytochrome *c* (3 μM), and then exposed to carbon monoxide. To initiate the reaction, the enzyme solution was mixed with oxygen-saturated (1.4 mM) buffer and the CO was photodissociated.

TABLE 1 Frequencies and isotope shifts (cm⁻¹) of isotope-sensitive stretching modes in intermediates of cytochrome *c* oxidase

Mode	Buffer	Frequency ¹⁶ O ₂ (¹⁸ O ₂)	Δ(¹⁶ O ₂ - ¹⁸ O ₂)	Δ(H-D) ¹⁶ O ₂ (¹⁸ O ₂)
Fe-O ₂	H ₂ O, D ₂ O	568 (547)	21	0
Fe ⁴⁺ =O ²⁻	H ₂ O	786 (751)	35	-15 (-12)
Fe ⁴⁺ =O ²⁻	D ₂ O	801 (763)	38	
Fe ³⁺ -OH ⁻	H ₂ O	450 (425)	25	8 (6)
Fe ³⁺ -OH ⁻	D ₂ O	442 (419)	23	

residue in the distal pocket. Moreover, in many cases when the ferryl-oxo complex is hydrogen bonded, suspension in deuterated buffers leads to an increase in frequency of the ferryl-oxo mode ($2\text{--}5\text{ cm}^{-1}$); this has been interpreted as either a coupling of the ferryl-oxo mode to a librational mode of a nearby water molecule (in cytochrome *c* peroxidase)¹⁵, or to a difference in hydrogen bond strength in D_2O compared with H_2O ¹³. The change in the ferryl-oxo mode of cytochrome *c* oxidase deuteration ($12\text{--}15\text{ cm}^{-1}$) is of the same order as that resulting from the complete loss of hydrogen bonding to the oxo atom in other haem proteins. Also, the frequency of the mode in D_2O is the same as that reported¹⁷ for myoglobin in H_2O , in which there is no evidence for hydrogen bonding. On the basis of these results, we conclude that the ferryl-oxo complex is hydrogen bonded in protonated buffer, and that conformational changes induced by deuteration remove the proton (deuteron) donor to result in a non-hydrogen-bonded ferryl species in D_2O . We postulate that the proton donor for the hydrogen bond to the ferryl-oxo complex is a water molecule loosely associated with Cu_B . A ferryl intermediate has previously been proposed to exist at the 3-electron O_2 reduction stage^{2,3,18–23}. Our data offer direct evidence for a ferryl intermediate and show that it is hydrogen bonded; they also provide an isolated marker line which should be useful for kinetic studies.

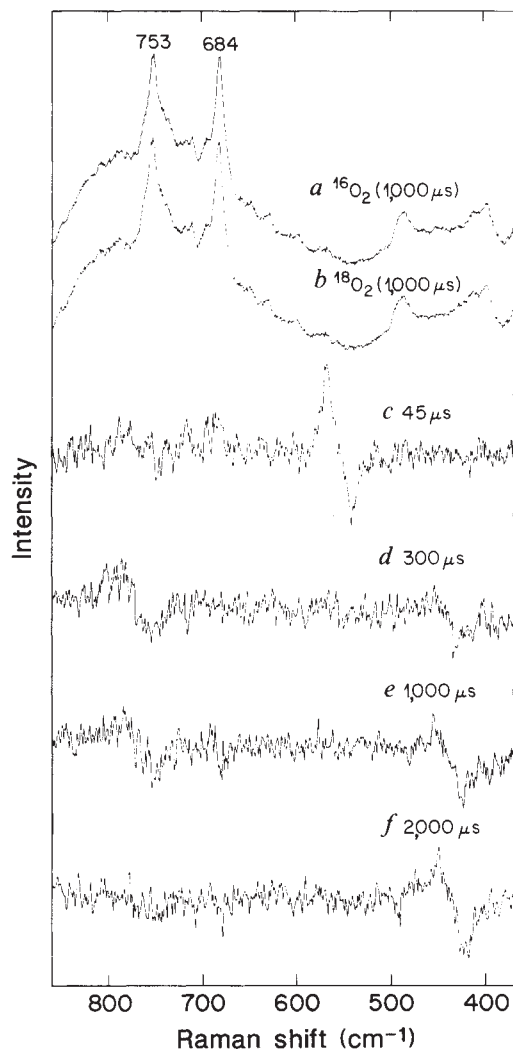


FIG. 2 Spectra of intermediates in reaction of $^{16}\text{O}_2$ (a) and $^{18}\text{O}_2$ (b) with cytochrome *c* oxidase at $100\text{ }\mu\text{s}$ after initiation of the reaction. c–f, Selected $^{16}\text{O}_2\text{--}^{18}\text{O}_2$ difference spectra. Only as a result of obtaining the well balanced difference spectra did the oxygen isotope-sensitive lines become apparent. Each difference spectrum is a result of subtracting spectra that were accumulated for 30 s .

TABLE 2 Frequencies and isotope shifts (cm^{-1}) for ferryl-oxo stretching modes in selected haem proteins

Protein	pH	Frequency $^{16}\text{O}_2$ ($^{18}\text{O}_2$)	$\Delta(^{16}\text{O}_2\text{--}^{18}\text{O}_2)$	$\Delta(\text{H--D})$	Refs
HRP (II)*	<8.5	775 (743)	32	−4	11–14
HRP (II)*	>8.5	788 (755)	33	0	11–14
CcP (ES)†	4–11	767 (727)	40	−5	15
Catalase (II)	7	775 (746)	29	−2	16
Myoglobin	8.6	797 (771)	26	0	17

* HRP (II) is compound II of horseradish peroxidase.

† CcP (ES) is compound ES of cytochrome *c* peroxidase.

As the ferryl species decays, a new intermediate with an isotope-sensitive line at 450 cm^{-1} is formed (Figs 1 and 2). Comparison of our observed frequency and isotope frequency shifts (Table 1) with those reported for other haem proteins^{24–27}, and with calculated values of frequency shifts, led us to assign this line as an $\text{Fe}^{3+}\text{--OH}^-$ stretching mode of an iron-hydroxide intermediate. The late appearance of this hydroxide intermediate and its survival to 3 ms indicate that it participates at the level of 4-electron reduction. This is consistent with the high-frequency resonance Raman spectrum¹⁰ (data not shown) in which cytochrome *a* is essentially fully oxidized. Iron-hydroxide intermediates have been proposed before^{2,3} at the 4-electron stage in the O_2 reduction process but until now have not been confirmed. The frequency of the $\text{Fe}^{3+}\text{--OH}^-$ mode is significantly lower than in most other hydroxides^{24–26} (for example, in metmyoglobin and methaemoglobin hydroxides, the $\text{Fe}\text{--OH}$ mode is near 490 cm^{-1}). Indeed, the $\text{Fe}\text{--OH}$ mode is lower than that of a hydroxide complex in cytochrome *c* oxidase (477 cm^{-1}) that we detected earlier under continual laser illumination of the ferric enzyme²⁷. We attribute this low frequency in the transient intermediate to a weak $\text{Fe}\text{--OH}$ bond which facilitates its eventual release.

With this identification of a ferryl and a hydroxy intermediate, the predominant intermediates at nearly each redox level are now established. The direct observation of a ferryl and a hydroxy intermediate at the levels of 3- and 4-electron reduction, respectively, supports the catalytic mechanism proposed by Wikström² and others³. Confirmation of the ferryl intermediate is particularly important in view of its postulated role in proton translocation². □

Received 8 June; accepted 29 August 1990.

- Oliveberg, M., Brzezinski, P. & Malmström, B. G. *Biochim. biophys. Acta* **977**, 322–328 (1989).
- Wikström, M. *Nature* **338**, 776–778 (1989).
- Chan, S. I. & Li, P. M. *Biochemistry* **29**, 1–12 (1990).
- Varotsis, C., Woodruff, W. H. & Babcock, G. T. *J. Am. chem. Soc.* **111**, 6439–6440 (1989); **112**, 1297 (1990).
- Han, S., Ching, Y.-C. & Rousseau, D. L. *Biochemistry* **29**, 1380–1384 (1990).
- Han, S., Ching, Y.-C. & Rousseau, D. L. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2491–2495 (1990).
- Argade, P. V., Ching, Y.-C. & Rousseau, D. L. *Science* **225**, 329–331 (1984).
- Han, S., Ching, Y.-C. & Rousseau, D. L. *J. Am. chem. Soc.* (in the press).
- Hill, B. C., Greenwood, C. & Nicholls, P. *Biochim. biophys. Acta* **853**, 91–113 (1986).
- Han, S., Ching, Y.-C. & Rousseau, D. L. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Terner, J., Sitter, A. J. & Reczek, C. M. *Biochim. biophys. Acta* **828**, 73–80 (1985).
- Sitter, A. J., Reczek, C. M. & Terner, J. *J. biol. Chem.* **260**, 7515–7522 (1985).
- Hashimoto, S., Tatsuno, Y. & Kitagawa, T. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2417–2421 (1986).
- Hashimoto, S., Nakajima, R., Yamazaki, I., Tatsuno, Y. & Kitagawa, T. *FEBS Lett.* **208**, 305–307 (1986).
- Hashimoto, S., Tareoka, J., Inubushi, T., Yonetani, T. & Kitagawa, T. *J. biol. Chem.* **261**, 11110–11118 (1986).
- Chuang, W.-J., Heldt, J. & VanWart, H. E. *J. biol. Chem.* **264**, 14209–14215 (1989).
- Sitter, A. J., Reczek, C. M. & Terner, J. *Biochim. biophys. Acta* **828**, 229–235 (1985).
- Wikström, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4051–4054 (1981).
- Wikström, M. *Chemica Scripta* **27B**, 53–58 (1987).
- Wikström, M. *Chemica Scripta* **28A**, 71–74 (1988).
- Blair, D. F., Witt, S. N. & Chan, S. I. *J. Am. chem. Soc.* **107**, 7389–7399 (1985).
- Witt, S. N., Blair, D. F. & Chan, S. I. *J. biol. Chem.* **261**, 8104–8107 (1986).
- Kumar, C., Naqui, A., Powers, L., Ching, Y.-C. & Chance, B. *J. biol. Chem.* **263**, 7159–7163 (1958).
- Asher, S. A., Vickery, L. E., Schuster, T. M. & Sauer, K. *Biochemistry* **16**, 5849–5856 (1977).
- Asher, S. A. & Schuster, T. M. *Biochemistry* **18**, 5377–5387 (1979).
- Desbois, A., Lutz, M. & Banerjee, R. *Biochemistry* **18**, 1510–1518 (1979).
- Han, S., Ching, Y.-C. & Rousseau, D. L. *J. biol. Chem.* **264**, 6604–6607 (1989).
- Yonetani, T. *J. biol. Chem.* **235**, 845–852 (1960).

ACKNOWLEDGEMENTS. S.H. is partially supported by the National Institute of General Medical Sciences.

The RNA polymerases from T7 and related bacteriophages, in conjunction with elements of DNA and RNA viruses, can be used in novel ways for expression of genes in mammalian cells.

91

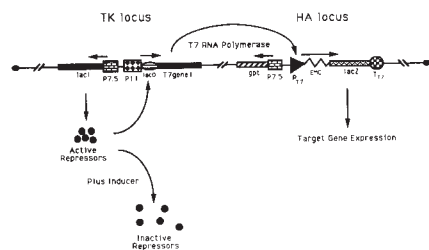


FIG. 2. Recombinant virus Vac/Op/T7, in which virus-encoded T7 RNA polymerase is induced by isopropyl β -D-thiogalactoside (IPTG). *LacI* is regulated by the vaccinia P7.5 early/late promoter and is therefore expressed throughout infection. Active repressor tetramers bind *lacO*, positioned between the vaccinia P11 late promoter and the T7 gene 1 which encodes T7 RNA polymerase. Under these conditions, expression of T7 RNA polymerase is suppressed. Upon addition of inducer IPTG to the medium, repressor is inactivated and expression of T7 RNA polymerase is induced. T7 RNA polymerase initiates transcription downstream of the T7 promoter (P_{T7}), which stops at the T7 terminator (T_{T7}). In the example, the *lacZ* gene is transcribed leading to expression of β -galactosidase. The *gpt* gene was used as dominant selectable marker in order to construct the recombinant vaccinia virus. HA and TK refer to the nonessential haemagglutinin and thymidine kinase genes used as sites for gene insertion.

could be expressed, provided the EMCV UTR was present at the 5' end of the mRNA. This system may be valuable for analysing structural features of mRNA that are important for stability and translation. The level of expression was not very high unless the cells were also infected with vaccinia virus. The role of the virus in this situation is unclear, particularly as capping is not required for translation of RNAs that have EMCV UTR sequences. The cell line provides a means for high-level expression when infected with a recombinant vaccinia virus containing the target gene regulated by a T7 promoter and EMCV UTR. Stable cell lines containing the gene for bacteriophage T3 RNA polymerase have been constructed and these lines also require vaccinia virus¹³ and/or the EMCV UTR¹⁴ for efficient expression.

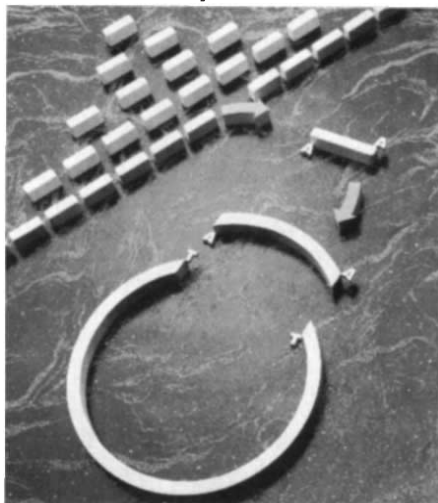
Uses and future prospects

The T7 vector system provides a rapid and powerful way of analysing the expressed products of cloned genes. Moreover, the system is integrated, in that the same plasmid can be used for *in vitro* transcription and translation in reticulocyte extracts, for analytical transfection assays in cells infected with recombinant vaccinia virus encoding T7 RNA polymerase, and for the construction of recombinant vaccinia viruses that allow protein production in preparative amounts. The use of vaccinia virus with the T7 system, however, has limitations. Vaccinia virus kills the cells and so expression is a batch process rather than a continuous one. In addition, the

Recombinant DNA technology

A kit for the direct cloning of PCR-amplified nucleic acids, a novel DNA sequencing technology and mRNA spin columns — new tools for the molecular biologist.

For the **direct cloning of PCR-amplified nucleic acids** without the use of restriction or modifying enzymes, Invitrogen recommends its TACloning kit (*Reader Service No. 101*). The TACloning procedure takes advantage of the terminal transferase activity inherent in thermo-



Invitrogen's TACloning kit for the direct cloning of PCR products.

philic polymerases, which results in the addition of a single 3' deoxyadenylate residue to each strand of amplified product. The prepared pCR2000 vector contained in the TACloning kit possesses single 3' thymidylate residues to allow direct ligation of amplified material. Each \$245 (US) kit contains sufficient reagents for 20 reactions.

virus, despite its historical use as a human vaccine, is infectious and so precautions need to be taken; use of highly attenuated or defective viruses would be desirable. Thus far, the system has been developed primarily for cytoplasmic transcription. However, transient expression has been achieved with phage T7 RNA polymerase that has been targeted to the cell nucleus¹⁵ and further work may lead to stable, inducible, high-level expression. □

Bernard Moss, Orna Elroy-Stein and Tamio Mizukami are from the Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. Thomas Fuerst and William Alexander are at MedImmune, Gaithersburg, Maryland 20878, USA. Present address of Tamio Mizukami is Tokyo Research Laboratories, Kyowa Hakko Kogyo Co., Ltd, Machidashi, Tokyo 194, Japan. For more information, fill in reader

Poly (NAT) ready-to-use gels from ELCHROM are designed to resolve DNA fragments of small-to-medium size, with optimal separation in the 200–2,000-bp range (*Reader Service No. 102*). The gels, which are supplied precast on plastic supports, are made of the monomer NAT (N-acryloyl-Tris) and a crosslinker. Poly (NAT) gels can be run in horizontal electrophoresis chambers and are particularly useful for the analysis of DNA restriction fragment length polymorphisms and DNA amplification products.

The latest optional accessory for the Gene Construction Kit, Textco's Macintosh-based software program for the graphic manipulation of DNA sequences with plasmid drawing capabilities, is the **file searching accessory module** (*Reader Service No. 103*). Using the module, construct files created with the Gene Construction Kit can be searched for matches with key words in comments or DNA sequences. Search criteria are typed into the dialogue box and, in one pass, all relevant sequences can be found. Up to 40 different search criteria can be entered, each linked by boolean operators (and, or, not). Searches can be carried out in the background using multifinder, while the user continues to work in other software applications. The file searching accessory module costs \$350 (US).

The TE 77 Semiphor semi-dry transfer unit from Hoefer is designed to transfer

service number 100.

1. Fuerst, T.R., Niles, E.G., Studier, F.W. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8122–8126 (1986).
2. Mizukami, T., Fuerst, T.R., Berger, E.A. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **85**, 9273–9277 (1988).
3. Whitt, M.A., Chong, L. & Rose, J.K. *J. Virol.* **63**, 3569–3578 (1989).
4. Pattnaik, A.K. & Wertz, G.W. *J. Virol.* **64**, 2948–2957 (1990).
5. Fuerst, T.R., Earl, P.L. & Moss, B. *Molec. cell. Biol.* **7**, 2538–2544 (1987).
6. Barrett, N. et al. *AIDS Res. Human Retroviruses* **5**, 159–171 (1989).
7. Fuerst, T.R. & Moss, B. *J. molec. Biol.* **206**, 333–348 (1989).
8. Elroy-Stein, O., Fuerst, T.R. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6126–6130 (1989).
9. Rich, D.P. et al. *Nature* **347**, 358–363 (1990).
10. Fuerst, T.R., Fernandez, M.P. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2549–2553 (1989).
11. Alexander, W.A., Moss, B. & Fuerst, T.R. (in preparation).
12. Elroy-Stein, O. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6743–6747 (1990).
13. Deuschle, U. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5400–5404 (1989).
14. Zhou, Y., Giordano, T.J., Durbin, R.K. & McAllister, W.T. *Molec. cell. Biol.* **10**, 4529–4537 (1990).
15. Lieber, A., Kiessling, U. & Strauss, M. *Nucleic Acids Res.* **17**, 8485–8493 (1989).

gels in about 60 minutes using low current and voltage and a minimum of buffer (Reader Service No. 104). The 21 × 26-cm screen electrodes — platinum anode and stainless steel cathode — are large enough to transfer several large gels at a time and



Hoefer's semi-dry gel transfer unit.

many more small ones. Membranes are free of electrode deposits after transfer, says Hoefer. The \$925 (US) TE 77 transfer unit is constructed of foam-moulded Noryl with safety interlocks, and is supplied with solid and pre-cut plastic masks and blotter paper.

DNA sequencing developments

In collaboration with New England BioLabs, Millipore has developed a novel **DNA sequencing technology** — the Multiplex system — that allows up to 10 DNA fragments to be sequenced simultaneously and which replaces radioactive detection with chemiluminescence (Reader Service No. 105). When using the Multiplex sequencing kit, multiple DNA fragments are tagged with 'Plex' tags, which are derived from genetically-engineered plasmids. Each plasmid contains pieces of DNA which act as the identifying tags by becoming incorporated into the fragments being sequenced. The fragments are then pooled and processed as if they were one. Sequencing reactions are performed as usual. After gel electrophoresis, the DNA is transferred and crosslinked onto a membrane. Each DNA sequence is identified by hybridization with the appropriate biotinylated probe and subsequent chemiluminescent detection. Bands can be detected in under 10 minutes, as compared to up to 72 hours when using standard radioactive isotopes, says Millipore.

According to AT Biochem, its Long Ranger DNA sequencing gel from the HydroLink family of gels can produce up to 30 per cent more information than equivalent acrylamide gels (Reader Service No. 106). Long Ranger is a chemically modified acrylamide with new monomers and crosslinkers for increased resolving power and mechanical strength. The gel is supplied as a 50 per cent concentrate and can be substituted for acrylamide-bis stock in sequencing pro-

ocols. Long Ranger costs \$100 (US) for 250 ml, which makes 2.5 litres of a 5 per cent gel.

For increased productivity in large-volume DNA sequencing laboratories, Beckman has introduced the Autoreader **automated film reader** which combines automatic film reading and data analysis (Reader Service No. 107). Autoreader



Autoreader speeds DNA sequence analysis. It takes 1.5 minutes to scan the DNA sequencing autoradiogram and about 15 minutes to generate sequence data from 10 templates, or a total of 40 lanes: it can take an expert up to 2 hours to gather the same amount of information by manual means, says Beckman. For overnight unattended operation, Autoreader can be loaded with up to 10 films.

Pure and simple

Poly A⁺ mRNA can be purified from total RNA in 45 minutes, says 5' → 3', when using its **mRNA spin columns** (Reader Service No. 108). A maximum of 1.5 mg of total RNA prepared from tissue or cultured cells can be applied to each oligo (dT) cellulose spin column. At least one per cent of the total RNA is obtained as poly A⁺ mRNA, says 5' → 3'. Columns can be purchased separately, costing \$120 (US) for a pair. Alternatively, the company offers a complete kit, costing \$170 (US), which includes two columns, RNase-free buffers and reagents. The columns can be reused for a total of 4 two-cycle or 8 one-cycle mRNA purifications.

A recent introduction by BIO 101 is the RNaid kit for the **extraction of RNA from agarose or polyacrylamide gels** (Reader Service No. 109). RNA can be isolated from both denaturing and non-denaturing gels. Typically, RNA purification takes 20



RNaid for RNA extractions with ease.

minutes, says BIO 101. The RNaid kit includes all the reagents necessary to carry out several experimental protocols. The user can size-select RNA before making a cDNA library, clean up RNA PCR products, remove unincorporated label from RNA probes to reduce background noise, desalt and rapidly remove other contaminants such as enzymes and proteins from RNA samples, and rapidly isolate total RNA from guanidinium cell lysates. Each RNaid kit costs \$94 (US) and contains sufficient RNase-free reagents for about 100–200 purifications.

StrataClean **resin**, a non-toxic alternative to small-volume phenol/chloroform

ADVERTISEMENTS

Synthetic Oligonucleotides

Custom made Probes and Primers,
Synthesised, purified and shipped in 3 days

£4.00 per base

(All inclusive)

0.2 µM scale (approx. 500 µg)

Modified bases, aminoalkylation,
data base analysis and full design facilities

Molecular Medicine Unit,
King's College School of Medicine & Dentistry,
Denmark Hill, London SE5 8RX

Tel. XX-44-(0)71 - 326 3126

FAX XX-44-(0)71 - 733 3877

Reader Service No.56

Custom DNA

Purified and Shipped in 48 hours,
Worldwide.

100 µg, \$10.00 a base for the first 15 bases,
\$7.50 for each additional base.
Delivered.

Research Genetics

1-800-533-4363

In the United Kingdom 0-800-89-1393

In Canada 1-800-533-4363

FAX 205-536-9016

Reader Service No.26

LIPOSOMES

REPRODUCIBLE — HOMOGENEOUS —
EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5–10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

BIANORM

POB 126, D-8000 Munich 65, FRG

Reader Service No.49

ADVERTISEMENTS

**CUSTOM-MADE & PURIFIED
DNA PEPTIDES**

\$8/base
50–100 µg
40 hour delivery

Quotes available:
50–300 mg
1–3 wks delivery

GUARANTEED QUALITY & QUANTITY

V.G. CORP.

1200 Wonderland Rd. S.
Bldg #7, Unit #2
London, Ontario
Canada N6L 1A8

CALL OR FAX: (519) 652-3758

Reader Service No.29

Serving now
BULK AMOUNTS
of

**90 RESTRICTION - and
20 MODIFYING ENZYMES**

We guarantee excellent quality
and very competitive prices.
Please inquire for quotations



AGS Angewandte Gentechnologie
Systeme GmbH
Riescherstraße 12, D-6900 Heidelberg 1
West-Germany
Tel.: Germany (6221) 81023
Fax: (6221) 840610

Reader Service No.80

**TIB MOLBIOL BERLIN
CUSTOM DNA SYNTHESIS**

DM per base
5,-
DM 95,- per oligo

FAX (0049) 30 6261960

Reader Service No.79



SINGER
INSTRUMENTS



**COMPLETE SYSTEMS FOR
XENOPUS
OOCYTE INJECTION**

INT. SALES: +44. 984. 40226.

FAX: + 44. 984. 41166

Singer Instrument Co. Ltd.

Roadwater, Watchet, Somerset,
England, TA23 0QL. Tel: Washford (0984) 40226.
Telex: 337492 Comcab G

Reader Service No.50

extractions of restriction and modifying enzymes from DNA, is available from Stratagene (*Reader Service No. 110*). The resin consists of hydroxylated silica particles, which exhibit similar characteristics to phenol but without the hazards, says Stratagene. It is supplied as a 25 per cent slurry, is non-toxic, non-flammable, non-carcinogenic, and does not require special disposal measures. At neutral pH, StrataClean resin has a high affinity for proteins and a very low affinity for nucleic acids (1,900:1, respectively). The resin has been shown to be as effective as phenol/chloroform for the removal of calf intestinal alkaline phosphatase, *Taq* DNA polymerase, exonuclease III and DNase I.

Molecular miscellany

The Australian biotechnology company, GroPep, offers two new **analogues of insulin-like growth factor-I** (IGF-I), which are produced by recombinant DNA technology and which are more potent than IGF-I in cell culture and in stimulating growth *in vivo* (*Reader Service No. 111*). Des(1–3)IGF-I, which lacks the N-terminal tripeptide of IGF-I, exhibits diminished binding to some classes of IGF binding proteins, says GroPep. Long R³ IGF-I, however, has a substitution at residue 3 (Arg for Glu) and a 13 residue N-terminal extension peptide, and exhibits diminished binding to all classes of IGF binding proteins. Both recombinant growth factors are available in 100-µg and 1-mg quantities as receptor-grade, freeze-dried preparations. Long R³ IGF-I is also available as a media-grade preparation in 100-µg, 1-mg and 5-mg quantities.

Molecular biology-grade **alkaline phosphatase**, a useful tool for the cloning and labelling of nucleic acids, is available from United States Biochemical in a form that has a high specific activity and is easily heat-inactivated (*Reader Service No. 112*). The enzyme is isolated from artichoke and has a similar specific activity to alkaline phosphatase isolated from calf intestine. However, unlike the latter, it is completely and irreversibly inactivated in Tris buffers at pH 8–8.5 by heating for 20 minutes at 65 °C. Thermolabile alkaline phosphatase is available in 500-, 1,000- and 5,000 units for \$25, \$42 and \$195 (US), respectively.

The InBed kit provides, says New England BioLabs, a convenient means of **embedding intact chromosomes from mammalian, plant or bacterial cells in agarose** (*Reader Service No. 113*). Protection of large DNAs from shear forces is critical for chromosomal mapping and fragment identification by pulsed-field gel electrophoresis, says NEB. The \$150 (US) Imbed kit is supplied with all the necessary reagents, including moulds for the GelSyringe system which is designed to

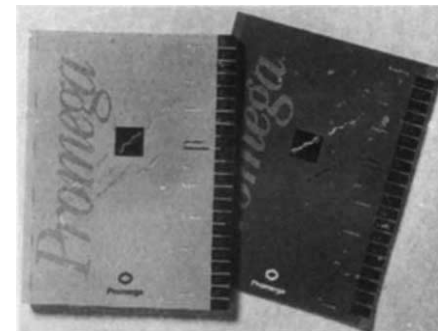


NEB's Imbed kit — a convenient way to produce agarose-embedded chromosomes.

provide a convenient means of dispensing agarose plugs that are of consistent size and quality.

Product guides

Hot off the press is the 1990/1991 **molecular biology catalogue** and price list from Promega (*Reader Service No. 114*). New products featured include purification kits based on paramagnetic particles and sequencing kits that use modified *Taq*



Product news for '90/91 from Promega.

polymerase. New research areas for Promega include cellular regulation, protein analysis, cell culture and eukaryotic gene expression/regulation.

The 4th edition of the free **ATCC/NIH repository catalogue of human and mouse DNA probes and libraries** is now available from the American Type Culture Collection (ATCC) (*Reader Service No. 115*). Listed in the 215-page catalogue are materials deposited at the ATCC as part of the genetic repository supported by the NIH, National Institute of Child Health and Human Development and the Division of Research Resources. Related materials from ATCC's own collection and Patent Culture Depository are also listed. Product lines include: human chromosome-specific genomic libraries, human and mouse genetic probes and cloned genes, probes and clones for oncogenes and transforming proteins, bacterial hosts for transformation or plating libraries, and other cDNA and genomic libraries. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.